

FORMAL AND INFORMAL SECTORS: IS THERE ANY WAGE DIFFERENTIAL?

ABSTRACT

The main objective of this paper is to investigate if a wage difference exists between formal and informal sectors in the case of the Turkish labour market using a sample of wagedworkers. To this end, we use data for 2004 and 2009 and a novel definition of the informal sector. On the methodological front, we adopt three alternative decomposition techniques, namely, the Oaxaca-Ransom (1994) decomposition in the context of mean regression, the Machado and Mata (2005) decomposition in the quantile regression framework and the non-parametric decomposition method proposed by Nopo (2008). The results reveal the existence of a wage gap between the two sectors. We found education and experience to be key determinants of earnings. The findings of this paper have implications for policies, which might be directed towards developing approaches with a focus on education and experience.

Keywords: Informal Economies; Labour Market; Wage; Turkey.

1. INTRODUCTION

The informal sector is a broad-ranging concept, which is difficult to define. Indeed, it generally includes a large number of subjects from legal to illegal and people from profit maximising entrepreneurs to street vendors, etc. Since it covers a wide range of economic activities, it is a multidimensional phenomenon encompassing a wide range of definitions and different measurement methods. Much theoretical and empirical research has been carried out relating to the informal sector and its definition and measure, reflecting its importance, showing that the informal sector is very difficult and challenging due to its mysterious nature and absence of reliable data over time (see Buehn and Schneider, 2013; Schneider and Williams, 2013; Williams, 2013). According to ILO (1991), the informal sector includes activities of the working poor who are not recognised, recorded, protected and regulated by public authorities. In contrast, Maloney (1999; 2004) argues that the informal sector cannot be defined as marginal and a disadvantaged sector. This is because it might be a voluntary choice for individuals and/or businesses due to various factors i.e. tax evasion, avoiding the cost of regulations, having greater flexibility and more freedom. Others, for example, Portes et al, (1989), define the informal sector as a process of income-generating activities that are not regulated by institutions of the society in a legal and social environment. Recently researchers use narrow informal sector definition, which can be defined as the paid informal sector (see Williams, 2012 and Vandenberg, 2014; among others). Williams (2011, p.25) defined informal sector as “monetised exchange that is unregistered by or hidden from the state for tax, social security and/or labour law purposes but which is legal in all other respects”. What emerges from these studies is that there is no universally accepted unique definition of what the informal sector is. Despite this disagreement, it is widely accepted that the informal sector plays a key role in the global economy (see Imamoglu et al, 2018; Williams and Nadin,

2012a, 2012b). Given this importance, it is crucial to try to gauge a better understanding of the informal sector. This paper explores two interrelated issues. Firstly, it investigates if there is a wage difference between the formal and informal sectors in Turkey and the extent of such difference. Secondly, the paper investigates the key factors explaining the wage differential between the two sectors.

The rest of the paper is structured as follows: Section 2 provides the literature review. In Section 3 we discuss the data and provide alternative definitions of labour market informality used in empirical analysis. Section 4 presents the econometric methodology, whilst Section 5 focuses on the discussion of our main findings. Finally, the last section concludes the paper.

2. LITERATURE REVIEW

A considerable number of studies have appeared in the literature, which focused on the informal economic activities (Williams, 2011a; 2010; 2008; Williams & Round, 2009; Imamoglu, 2016; Fethi et al., 2006; 2004). On the other hand, economists have studied the formal-informal sector wage gap attempting to explore whether informal workers earn less than their formal counterpart. Alternative theories have been put forward in an attempt to explain this phenomenon. Along this, the traditional labour theory (see Harris and Todaro, 1970; Field, 1990) argues that workers choose the informal sector because they are rationed out of the formal sector due to entry barriers (e.g. labour market regulations). Consequently, workers settle for informal sector jobs with lower wages. On the other hand, according to the competitive labour market theory (see Maloney, 1999 and 2004; Perry et al. 2007; Henley et al. 2009) working in the informal sector reflects the worker's own choice, which is based on the advantages and disadvantages associated with each sector. Under this framework, the

wage gap tends to disappear. The formal-informal wage gap is crucial to understand labour market informality, especially in developing countries, which are characterised by large informal sector. The reported evidence, however, appears to be mixed and inconclusive. Indeed, whilst some studies find that workers in the informal sector are paid less than their formal counterpart, other studies show that wage differentials between the two sectors may not be a stylised fact. For example, Marcouiller et al. (1997) show a significant wage premium in the formal sector in El Salvador and Peru, while in Mexico informal workers have higher wage premium. Monsted (2000) finds a significant formal-informal wage inequality for Bolivia. Henley (2001), in the context of Brazil, found that informal workers tend to earn less on average than their formal counterparts. Tannuri and Pianto (2002) confirm a positive wage gap in favour of formal workers throughout the entire wage distribution in Brazil. Bargain and Kwenda (2009) also measure the formal-informal wage gap in Brazil, South Africa, and Mexico. Their results indicate a significant gap at lower quantiles across all countries studied. Sookram and Watson (2008) show that formal sector workers earn more than informal sector workers and work experience is the most important determinant of the wage gap in Caribbean islands Trinidad and Tobago.

Sample selection bias, which may result from either self-selection of individuals into different employment types (formal or informal), is the challenge in investigating the wage differential. Recently, some studies have focused on matching approaches to deal with a possible sample selection problem. For example, Pratap and Quintin (2006) find an insignificant wage gap in Argentina using the propensity score matching method (PSM) in contrast to OLS on standard Mincerian wage equations, which showed a significant formal-informal wage differential. Badaoui et al. (2008) examine whether individuals working in the informal sector suffer a wage penalty in South Africa. Using OLS regression, they find a significant informal wage

penalty. However, using the PSM method the results show a substantial decrease in the estimated wage penalty. Daza and Gamboa (2013) measure the wage gap between the two sectors in Colombia and find that formal workers earn between 30% and 60% more, on average than informal workers. In contrast to Colombia, in a more recent paper, Staneva and Arabsheibani (2014) find a high level of informal employment in Tajikistan with a significant wage gap in favour of informal workers.

3. WHY SHOULD INFORMAL SECTOR MATTER IN TURKEY?

Our focus in this paper is Turkey, which can be considered an interesting country. A number of studies have reported that Turkey is characterised by a large informal sector and is also marked by low labour force participation (Aydin et al. 2010; World Bank, 2010; Schneider et al. 2010; Ercan, 2011; Eurofound, 2013; TUIK, 2013; Simsek, 2014). The informal sector is a persistent and pervasive feature of the Turkish economy. Indeed, the informal sector is part of many people's daily routines in Turkey. Aydin et al. (2010), for example, argue that the informal sector is growing because of both supply and demand factors in Turkey. On the demand side, the integration of Turkey into global markets increases price competition and makes producers to minimise the cost of production which lead to demand-led growth in the informal sector. On the supply side, the authors argue that widespread rural-urban migration in Turkey, associated with privatisation and a decrease in public sector employment encourages some workers to accept jobs with lower wages and dire economic conditions in the informal sector. A report by Turkey's Ministry of Development in 2012 shows that with 24 January 1980 decisions, Turkey's integration to the world economy has increased and Turkey's economy has been undergoing structural changes from agriculture to industry. Labour is moving out of agriculture and rural areas, which is followed by a high level of unemployment, urban informality, high inflation, a high public-sector deficit which are the

main characteristics of the Turkish economy. These are followed by a series of domestic and global crises in the 1990s and 2000s and adversely affected the fragile Turkish economy. According to ILO (2012), Turkey suffers from unemployment problems of particularly young and educated people, due to job inadequacy in the formal sector. Job creation does not keep up with demographic challenges and the population growing faster than the job creation (see, Atiyas et al. 2017; Heper and Sayari, 2013; World Bank, 2010). In Turkey, a substantial amount of young people is struggling to find permanent formal jobs. The ILO highlights that informality among these young workers in Turkey is significant (58.8%) compared with adults (43.8%). The report states that it is because the formal sector is unable to create sufficient formal employment opportunities. According to the 2009 ILO Household Labour Force Survey, 60% of non-agricultural informal workers are wage employees and 30.7% are employers and own-account workers. The World Bank's (2010) report indicates that 55% of all jobs in construction, 39.6% in trade and 17% in services are informal. In a similar vein, Gursel (2014) points out that 85% of informal wage earners are employed in small firms. Simsek (2014), A Deputy Prime Minister, claims the informal sector is a key issue to tackle in Turkey because it slows the economic growth, production, cause significant loss of tax revenue and creates wage inequality. It is also important to note that Turkey has a higher degree of informality in comparison with European countries. For instance, the size of the informal sector is larger in Turkey than in the EU-27 countries (Buehn and Schneider, 2012).

The evidence presented suggests that the informal sector plays a very important role in Turkey's economy and has important economic and social functions. Therefore, measuring the size of the wage gap accurately between the formal and the informal sector represents an important aspect of labour market analysis to understand the functions of the Turkish labour market. Furthermore, it is worth examining the formal-informal wage gap in Turkey for

policy reasons. If the wage gap exists in favour of either formal workers or informal workers between the two sectors, this may suggest the existence of problems in the labour market and the need for policy action to prevent wage inequality as indicated by The Deputy Prime Minister. Wrong employment and social policies can cause undesired effects for people in the informal sector and can create an adverse effect on the economic performance of the country.

In addition, having reviewed the existing studies several important gaps in the literature are observed and this paper attempt to address them. First, there has been a little discussion about the wage gap between the formal and the informal sector and its implications in the Turkish labour market. The existing evidence is not sufficient to draw any reliable conclusion about informality, which is very important in designing policies. Second, although the formal-informal sector wage differential has been analysed in the context of Turkey (see for example, Tansel, 1999 and 2000; Baskaya and Hulagu, 2011; Tansel and Kan, 2012) most of these studies, with the exception of Tansel and Kan (2012), have used conventional regression analysis, implicitly focusing on the wage differential at the mean; thereby overlooking the distributional aspect.¹ Ignoring the wage gap between the wage distributions has strong limitations in explaining the formal-informal wage gap. Third, the existing studies have generally used the enterprise-based definition (firm size) or legalistic definition (social security status) in order to define labour market informality.² In her analysis of wage gap, Tansel (2000) define wage earners who are covered by social security institutions as formal and the self-employed have been regarded as informal. Baskaya and Hulagu (2011) used the semi-parametric technique to estimate the wage gap between the formal and the informal sector where informality is defined with respect to the registration status of individuals to

¹ Tansel and Kan (2012) use OLS and quantile regression analyses (including the informal sector dummy) and estimate the fixed effect OLS and fixed effect quantile regression.

² Appendix Table 1 provides a summary studies for the formal-informal wage gap reviewed in Turkey.

social security system excluding unpaid family workers, self-employed individuals and employers. In 2012, Tansel and Kan reported that formal wage earners earn significantly higher than informal wage earners and self-employed individuals. Using the quantile regression, they find a decrease in informal wage penalty with the earning levels. Kan and Tansel (2014) argue social security registration is a better measure of informality than enterprised based definition i.e. being employed in the small firm. In a more recent study, Baslevant and Acar (2015) define informal employment based on whether workers are registered with social security institution.

This paper extends the existing research in two ways. First, it aims to give a comprehensive formal-informal wage gap analysis in Turkey by adopting three different methodologies in the context of mean, quantile regression and the non-parametric decomposition method. It has been argued that the wage gap analysis should deal with three important issues; (i) common support problem; (ii) identification of the wage gap at different quantiles of the wage distribution; (iii) selection bias. The matching method does match individuals with comparable characteristics and they do not require an estimation of wage equations, arbitrary exclusion restrictions, which are required in the regression-based decomposition methods to solve sample selection bias. However, it must be stressed that these methods are not “immune” to selection bias (Nopo, 2008; Goraus, 2013; Van Der Velde et al. 2013). Second, there is no unique definition of the informal sector. It is therefore very difficult to find a most appropriate measure of the informal sector. There are alternative definitions of labour market informality i.e. social security registration, contract status, firm size. However, as Henley et al. (2006) argue, existing alternative definitions are not consistent and, thereby, the conclusion drawn from any of them will be subject to missing information from any other. The current paper, given the importance of understanding the wage gap between the two sectors and

nature of informality, adopts “novel” definitions of the informal sector, namely a lower-bound definition and an upper-bound definition; which are likely to capture some features of the lower-tier segment and an upper-tier segment of the informal sector. This, then, may provide further insights into designing appropriate policy actions.

4. DATA AND THE DEFINITION OF LABOUR MARKET INFORMALITY

4.1 Data

This paper uses two representative Household Budget Surveys (HBS) conducted by the Turkish Statistical Institute (TURKSTAT) and has been previously used by several studies on the Turkish labour market (e.g. Cudeville and Gurbuzer, 2007; Ben Salem and Bensidoun, 2012). HBS covers all settlements in Turkey and is specially designed to provide information on income and expenditure of the households. It asks information on age, marital status, education, employment status and income of individuals (regular and casual employees). In addition, they provide information on the socio-economic characteristics of the household, i.e. ownership status of land, house, and consumption expenditure patterns of the households. The two-stage stratified sampling method was used to select sample households. It, therefore, provides good quality and reliable data. Our focus in this paper is to investigate the wage gap between the formal and the informal sector of the Turkish labour market. Therefore, this study is based on an adult sample of the working-age population. Indeed, the sample includes individuals between the ages of 15 and 64 years, inclusively, to ensure that we do not include people who are likely to be in child labour or retired.³ The empirical analysis focuses on wage earners and we consider only the main job that individuals held on the survey month. Setting

³Turkish law establishes the minimum working age as 15 years whilst the retirement age is 65.

all restrictions and dropping all observations for which missing values are observed, we have a total of 5809 wage earners between the age 15-64 (4530 male and 1279 female) in 2004 and 7327 wage earners (5402 male and 1925 female). The results are reported for the sub-sample of wage earners broken down by the formal and the informal sector. The dependent variable is the logarithm of the hourly income. Hourly income is calculated by multiplying usual hours of work per week by 4.3 and dividing monthly income by imputed monthly hours of work. The monthly income is net income paid to people. Thus it includes transfers such as pension, social insurance deductions and taxes as well as extra income like bonuses, overtime, premiums etc. Hourly income is defined based on the assumption that individuals employed during the reference period are employed whole month. Thus, it is possible to identify individuals who were on job whole month. Weekly working hours are based on the main job. It includes overtime, but it does not include the lunch break and people commute to work. Those working under 35 hours a week are considered as part-time, whilst those who work more than 35 hours are classified as full-time.

Before going further, it should be noted that, although, Nopo's matching method (which is described later) eliminates the problem of specification of wage equations and thereby helping to prevent incorrect inferences we are aware that it does not eliminate the problem of choice variables. In this paper, our choice of variables is based on the human capital theory (Mincer, 1974) where education and experience are considered to be the most important determinant of wages. We further include variables that show personal characteristics i.e. marital status, sex and job characteristics with industry sector and geographical characteristics i.e. rural-urban that are found to be important in explaining the wage gap (Altonji and Blank, 1999). Overall, we believe that these variables play a key role in explaining the formal-informal wage gap precisely and cannot be ignored in our matching procedure. The determinants of wages

include a set of variables. The independent variables are listed as follows: (i) work experience, which is calculated as workers' age-adjusted by the years of schooling and school starting age.⁴ (ii) Sex, (iii) education, we use six dummies of education to capture the level of educational attainment of wage-earners, which are no education, primary school, primary education, high school, vocational high school and university. The base variable is no education.⁵ (iv) Marital status, three dummies for marital status are constructed. These are single, married, and another marital status. The base variable is single. (v) Employment location as rural or urban. (vi) We construct six dummies to capture the industrial sector. These are agriculture and mining sector, manufacturing and energy sector, construction sector, trade sector, transportation sector and finance and service sector. The base variable is the finance and service sector.

Table 1 portrays the composition of the lower-bound and upper-bound informality by age. It is apparent that both lower-bound and upper-bound informality is made of young people, between age 15-24 and 25-34. This may be explained by the fact that they may be rationed out of the formal sector, particularly, workers under lower-bound informality due to insufficient job creation in the formal sector as reported by ILO (2012) and Atiyas et al. (2017). Another possible explanation for this could be that being in the informal sector might be a decision given its flexible work hours, worker preferences, compensation differentials etc., in particular under the upper-bound informality. It is somewhat not surprising that informality is widespread among middle-aged workers between age 35 and 46. A possible explanation for this might be that generous retirement scheme in Turkey encourage workers to leave the labour market at a young age and get involved in informal economic activities

⁴ The school start age in Turkey is 6 years.

⁵ Primary school covers education between ages 6-11 and last 5 years. Secondary education includes all general, vocational and technical education after primary school and last 3 years. In 1997, 5-year compulsory education was expanded to 8-year, which consist of 5-year primary schools and 3-year secondary schools, which is so-called primary education.

because they are not required to pay taxes on their pension but are entitled to full health insurance.

“Place Table 1 about here”

4.1.1 The Definition of Labour Market Informality

There are alternative definitions of informality, which correspond to different ways of understanding the informal sector. These definitions are generally classified into three groups: productive definition, legalistic definition and the enterprise-based definition. In this study, the following definitions are used.

4.1.1.1 Definition I (enterprise-based definition): Firm size

A worker is defined as working in the informal sector if he/she works in firms employing less than 10 workers, with the exception of those working in the professional groups.⁶ Those who work in firms employing 10 or more employees are categorised as formal. We only apply this classification to the workers' main job. In this study, HBS's questions relating to firm size cover a number of workers in the workplace between less than 10, 10-24 workers, 25-49 workers and 50 workers and more. Hence, our smallest firm size is 10 or fewer workers.⁷

⁶Grouping of occupations are made according to ISCO 1988 (International Standard Classification of occupation). Occupational groups 1, 2 and 3, professionals, associate professionals and legislators respectively are excluded.

⁷According to World Bank (2010), informality tends to be widespread among small firms i.e. employing less than 10 workers in Turkey

4.1.1.2 Definition II (productive definition): Contract status

Individuals are classified as informal if they do not have a contract in their main job. There is no direct question whether individuals have a contract or not but surveys contain information on job permanency, individuals are questioned whether they are permanently employed, temporarily employed or they are occasional workers. Individuals who have a permanent job are classified as contract workers and formal but individuals who have temporary or occasional jobs are classified as non-contract workers and therefore informal.

4.1.1.3 Definition III: (legalistic definition) Social security status

Individuals are classified as informal workers if they are not registered by any institutions of social security in their main job and they are classified as formal workers if they are. HBS asks all individuals in employment aged 15 years and over, regardless of their employment status, whether they are registered by any social security related to their main jobs. Social Security registration includes social insurance institutions (SSI), the retirement fund (RF), Bag-Kur and private registration i.e. banks.

After identification of these three definitions, we construct two different estimates of informality, an upper-bound definition and a lower-bound definition of informality. The upper-bound informality represents the maximum potential number of workers in the informal sector according to our definitions. This includes those who are not registered for any social security or do not have a contract or are employed in a small firm. The lower-bound informality is the minimum potential number of workers and covers those who are not covered by any social security and do not have a contract and are employed in the small firm.

Unlike traditional and competitive approaches, some authors (e.g. fields, 1990) argue that the informal sector consists of “upper-tier segment” (voluntary entry exist) and “lower-tier segment”. Hence, it might be argued that lower-bound informality represents a lower-tier informal sector since workers in this category have no social security coverage, no contract and work in small firms. Upper-bound informality, on the other hand, can be considered as part of the upper-tier informal sector since it may include some workers close to formal workers. Table 2 shows the distribution of wage earners by informality status. It is very clear that workers can be employed without an employment contract, social security coverage and in small firms.

A full description of the variables, along with their meaning, is presented in the Appendix, Table 2. Descriptive statistics of the variables for the lower-bound and the upper-bound definition of informality are also reported in Table 3 in the Appendix. The 2004 sample consisted of 5014 formal workers and 705 informal workers. The 2009 sample consisted of 6498 formal workers and 829 informal workers. The descriptive analysis shows that formal workers, on average, earn more than informal workers and a significant proportion of formal wage earners are male under each definition of informality. Turning to education, we can see that a higher proportion of workers tend to work in the formal sector compared to the informal sector (as education levels increase). For the two samples, a higher proportion of formal workers have a university degree. However, primary school graduates dominate the informal sector. Looking at the industrial sector, we can see that agriculture and construction sectors dominate the informal sector while formal employment is more common in the finance and service sector under the upper bound and the lower bound informality. As distinct from a lower bound informality, a substantial number of workers in the trade sector are observed as informal under the upper bound informality.

“Place Table 2 about here”

5. METHODOLOGY

As stated earlier, we start the empirical analysis with the OLS regression and the decomposition in this framework i.e. the Oaxaca-Ransom (1994) decomposition. This is followed by the quantile decomposition technique based on the Machado and Mata (2005) technique and finally, we present the methodology based on a matching technique proposed by Nopo (2008).

5.1 OLS Regression and the Oaxaca-Ransom (1994) Decomposition

The analysis is based on a Mincerian earnings function, which is modelled by Mincer (1974) and includes experience (as a proxy for the job training), its square, education and other variables that are associated with individuals' earnings. This model is specified for the formal and the informal sector as follows:

$$\ln w_i^F = x_i^F \beta + \varepsilon_i^F \quad (1)$$

$$\ln w_i^I = x_i^I \beta + \varepsilon_i^I \quad (2)$$

where $i = \{1, \dots, N\}$ denotes individuals and $\ln w_i^F$ $\ln w_i^I$ refer to the log of the hourly wage of the individual i in the formal and the informal sectors, respectively. x_i^F and x_i^I represent

individual characteristics such as education, experience, gender, location etc. The estimated earning equations for formal and informal workers are used to calculate the difference in wages between the formal sector and the informal sector. The Oaxaca-Ransom (1994) decomposition can be written as follows:⁸

$$\ln \bar{w}_i^F - \ln \bar{w}_i^I = (\bar{x}^F - \bar{x}^I)\beta^* + \bar{x}^F(\hat{\beta}^F - \beta^*) + \bar{x}^I(\beta^* - \hat{\beta}^I) \quad (3)$$

The first term on the right-hand side indicates a wage gap due to endowment differences, the second term shows the difference between the actual and pooled returns for the formal sector and the third term captures the difference between actual and pooled returns for the informal sector. The main problem in measuring the wage gap between the formal and the informal sector is trying to answer the following counterfactual question: “What happens if the allocation of individuals to the informal sector and the formal sector are not random?” A participation decision of workers to participate in any sector (endogenous selection) brings selectivity bias to occur and affects the formal-informal wage gap. In the context of the mean regression literature, Heckman (1979) provides a solution to this issue and many studies have adopted this procedure to correct for sample selection. To implement this procedure, a probit regression is run in the first stage to find the probability of working in the informal sector as a function of original variables and an additional identifying variable. The key identifying variable in this analysis is the proportion of the informal sector workers to the number of all workers in each household. This identifier was used by Henley et al. (2009) for Brazil and is expected to have a positive impact on being informal, but it is not assumed to have any impact on wages. It is useful to note that measuring the labour market informality is subject to data availability. This then leads to a practical difficulty in finding excluded variables for

⁸ $\beta^* = \Omega \hat{\beta}^F + (1 - \Omega) \hat{\beta}^I$.

regression analysis. As mentioned above, workers may not be a random sample of working age population. Unobserved factors may determine workers' sectoral choice. We aim to solve the problem, sample selection, by introducing a correction term obtained from a probit model into the wage equation. Using the available data, we expect the proportion of the informal sector to all workers to influence the sectoral choice of workers without having a (direct) impact on wages. However, we are aware of potential problems related to the identification issue. The following selection equation is estimated using the probit separately for formal and informal workers.

$$I_i^* = \delta Z_i + \varepsilon_i \quad (4)$$

Where $I_i^* > 0$ shows workers are in the formal sector and $I_i^* \leq 0$ denote workers are in the informal sector. Z_i is a vector of individual characteristics that determine participation status and δ is coefficients of parameters. ε is the error terms distributed as $N(0, \sigma)$. In this process, the selectivity correction terms (λ) (Inverse Mills Ratio, IMR) is constructed for each sector. In the second stage, λ is included in wage equations as an additional regressor to control for potential selection bias and wage equations are estimated as follows:

$$\ln w_i^F = x_i^F \beta + \lambda^F \theta^F + \varepsilon_i^F \quad (5)$$

$$\ln w_i^I = x_i^I \beta + \lambda^I \theta^I + \varepsilon_i^I \quad (6)$$

Where, θ^F and θ^I are estimated coefficients of λ for formal and informal workers respectively.

In the presence of sample selection, we can then write the Oaxaca-Ransom decomposition as follows:

$$\ln \bar{w}_i^F - \ln \bar{w}_i^I = (\bar{x}^F - \bar{x}^I)\beta^* + \bar{x}^F(\hat{\beta}^F - \beta^*) + \bar{x}^I(\beta^* - \hat{\beta}^I) + (\hat{\theta}^F \hat{\lambda}^F - \hat{\theta}^I \hat{\lambda}^I) \quad (7)$$

The last term on the right-hand side represents the selectivity effect of the wage gap, which is attributable to selection bias. The selectivity effect is netted out from the overall wage gap. When the selection effect is netted out from the overall wage gap, the wage gap is called “offered wage gap” (Reimers, 1983; Neumann and Oaxaca, 2004; 2005; Boymond et al. 1994).

5.2 Quantile Regression and The Machado and Mata (2005) Decomposition

In this section, we apply the quantile regression analysis and the (MM) decomposition technique in this framework. One major advantage of the quantile regression over classical mean regression is that it captures workers’ heterogeneity in returns to observed characteristics and allows us to estimate the wage gap at different quantiles of the wage distribution. The quantile regression equations for formal and informal workers can be expressed as follows:

$$\ln w_{\theta i}^F = x_i^F \beta_{\theta} + \varepsilon_{\theta i}^F \text{ with } Q_{\theta}(\ln w_i^F | x_i^F) = x_i^F \beta_{\theta} \quad (8)$$

$$\ln w_{\theta i}^I = x_i^I \beta_{\theta} + \varepsilon_{\theta i}^I \text{ with } Q_{\theta}(\ln w_i^I | x_i^I) = x_i^I \beta_{\theta} \quad (9)$$

Where $\theta \in [0,1]$ and as before, $\ln w_i^F$ and $\ln w_i^I$ represent the log of hourly wages for the formal sector and the informal sector respectively for individuals $i = 1 \dots n$. $Q_\theta(\ln w_i | x_i)$ denotes the conditional quantile θ of $\ln w_i$. β_θ (slope of the quantile equation) in both QR wage equations represents the effects of changes in the set of individual and job characteristics on the θ^{th} conditional quantile of w_i . It is assumed that all quantiles of $\ln w_i^F$ and $\ln w_i^I$, conditional on x_i^F and, are linear in x_i^F and x_i^I . This then allows conditional quantiles of $\ln w_i^F$ and $\ln w_i^I$ to be estimated by linear quantile regression for each specific quantile.

The MM decomposition technique is based on simulation of the conditional marginal wage distribution estimated through linear quantile regression. The MM method is much more flexible than other decomposition methods, which extends the conventional Oaxaca wage decomposition on the mean to the entire wage distribution and combines a quantile regression and a bootstrap approach. The MM process to decompose the wage gap can be described as follows: First, in order to undertake the decomposition analysis, wage equations for each sector, given are as $\ln w_{\theta i}^F = x_i^F \beta_\theta$ and $\ln w_{\theta i}^I = x_i^I \beta_\theta$ for the formal sector and the informal sector respectively, are estimated across different quantiles of the wage distribution. Second, counterfactual wage distribution using estimates from the conditional quantile regression are constructed based on simulation technique following several steps described as below:

1. generate random sample size m ($m=5000$) from θ a uniform distribution in $[0,1]: \theta_1, \dots, \theta_m$

2. For each θ_m , estimate corresponding QR coefficients for the formal sector and the informal sector. $\hat{\beta}_{\theta_i}^F(\theta)$ and $\hat{\beta}_{\theta_i}^I(\theta) i = 1 \dots m$
3. Obtain a random sample of size m (m=5000) with replacement from the covariates x (characteristics of formal and the informal workers which were used to estimate QR coefficients) for each group of workers, $\{x_i^F\}_{i=1}^m, \{x_i^I\}_{i=1}^m$. The vector of characteristics for formal and informal workers is then used to estimate the log hourly wages for formal workers and informal workers, $\{ln\tilde{w}_{\theta}^F = \tilde{x}_i^F \hat{\beta}_{\theta}^F\}_{i=1}^m$ and $\{ln\tilde{w}_{\theta}^I = \tilde{x}_i^I \hat{\beta}_{\theta}^I\}_{i=1}^m$.
4. Construct a random sample of the counterfactual earnings distribution. That is: counterfactual earning distribution $lnw_i^{cf} = x^I \hat{\beta}_{\theta_i}^F$ shows the log of earning if informal workers had their own characteristics but were paid as formal workers.
5. Finally, using generated coefficients and characteristics, the wage gap is decomposed at different quantiles of the constructed earning distributions.

Machado and Mata (2005) argue that the above process produces a consistent estimator of counterfactual earning distribution and allows a comparison with the actual earning distribution. Therefore, the difference in earnings between the formal and the informal sectors at different quantiles are given as follows:

$$Q_{\theta}(lnw^F) - Q_{\theta}(lnw^I) = [Q_{\theta}(ln\tilde{w}_{\theta}^F) - Q_{\theta}(lnw_i^{cf})] + [Q_{\theta}(lnw_i^{cf}) - Q_{\theta}(ln\tilde{w}_{\theta}^I)] + resid$$

$$Q_{\theta}(lnw^F) - Q_{\theta}(lnw^I) = [Q_{\theta}(\tilde{x}_i^F \hat{\beta}_{\theta}^F) - Q_{\theta}(x^I \hat{\beta}_{\theta_i}^F)] + [Q_{\theta}(x^I \hat{\beta}_{\theta_i}^F) - Q_{\theta}(\tilde{x}_i^I \hat{\beta}_{\theta}^I)] + resid$$

(10)

The first term on the right-hand side is the characteristics components, which show the contribution of the differences in labour market characteristics between the θ th quantile of the formal sector wage distribution and the θ th quantile of the informal sector wage distribution. The second term on the right-hand side indicates the wage differential due to differences in coefficients. The residual term includes simulation errors and tends to disappear with more simulations.

Bargain and Kwenda (2013) argue that decomposition of the wage gap is not straightforward in the context of quantile regression. The authors state that there is no consensus regarding the most appropriate correction procedure for selectivity in this context. Buchinsky (1998) proposed the use of selection equation and approximation of IMR in the quantile regression. However, Buchinsky's estimation procedure is criticised on the basis that it suffers lack of robustness and it is very sensitive to identification; since it requires the independence assumption to obtain partially linear representation for the conditional quantile functions (Montenegro, 2001; Hyder and Reilly, 2005; Huber and Melly, 2011; Bargain and Kwenda, 2013). Buchinsky's approach relies on the assumption of conditional independence between the error terms and the regressors given the selection probability. This then implies that all quantile regression curves are parallel, and thereby limits the usefulness of quantile regression and sample selection correction in this framework.⁹ Another problem with the quantile sample selection model (using the MM decomposition method) is that it is computationally high demanding and very time-consuming. Many authors (i.e. Mizala et al. 2011; Goraus, 2013; Van Der Velde et al. 2013), now, apply an innovative approach proposed by Nopo (2008), which uses the matching technique to investigate the wage differential. This technique is considered as a potential solution to the problem of sample selection.

⁹For further discussion, see Huber and Melly (2011).

5.3 Nopo's (2008) Matching-Based Decomposition

In this section, we decompose the wage differential between the two sectors by applying the non-parametric decomposition proposed by Nopo (2008). The main aim of this matching technique is to compare different groups with similar characteristics. More precisely, Nopo's method uses matching procedure and selects sub-samples of formal and informal workers such that there are no differences in observable characteristics between the matched groups. Regression-based decompositions fail to capture differences in the support by estimation of wage equations for all formal and informal workers without restricting the comparison only to individuals with comparable characteristics (Barsky et al. 2002; Black et al. 2008; Nopo, 2008). The matching technique, especially the Nopo (2008), recognises the importance of "common support" (i.e. comparing formal and informal workers with similar observable characteristics). It also accounts for "out-of common support", which shows deviations for each group from the common support. Van Der Velde et al. (2013) argue that this matching technique has an advantage over conventional regression analysis. They contend that this approach does not only address the common support problem, it, also, allows the identification of the wage gap at different quantiles of the wage distribution. Furthermore, the authors state that this technique does not require an estimation of wage equations and arbitrary exclusion restrictions, which are required in the regression-based decomposition methods to solve the sample selection bias. Hence, it is a potential solution which is an alternative to the selection bias.

In this framework, the formal-informal wage differential can be decomposed as follows. Let W denote the random variable that models both formal and informal workers' earnings and x is the vector of individual characteristics i.e. education and marital status. Then, let the

functions $F^F(.)$ and $F^I(.)$ denote conditional cumulative distributions of individual characteristics, depending on being formal and informal respectively. Thus, $dF^F(.)$ and $dF^I(.)$ can be interpreted as the corresponding probability measures. Further, assume that $E[W/F, x] = g^F(x)$ is the expected value of earnings of workers in the formal sector conditional on individual characteristics and being formal. Similarly, $E[W/I, x] = g^I(x)$ is the expected value of earnings of workers in the informal sector, conditional on individual characteristics and being informal. Finally, let S^F and S^I indicate support for the distribution of individual characteristics for formal workers and informal workers respectively. In this framework, the formal-informal wage gap be derived as:

$$\Delta \equiv E[W/F, x] - E[W/I, x] \quad (11)$$

where Δ denotes the wage gap. Equation (11) can further be expressed as:

$$\Delta = \int_{S^F} g^F(x) dF^F(x) - \int_{S^I} g^I(x) dF^I(x) \quad (12)$$

Considering that support for the distribution of characteristics for formal workers, S^F , and support of the distribution of characteristics for informal workers, S^I , are different, each integral is then divided into two sub-groups: within the common support, $(S^F \cap S^I)$, and out of common support, $(S^I \cap \overline{S^F} \text{ and } \overline{S^I} \cap S^F)$. Equation (12) is then written as follows:

$$\left[\int_{S^I \cap S^F} g^F(x) dF^F(x) + \int_{S^F \cap \overline{S^I}} g^F(x) dF^F(x) \right] - \left[\int_{S^F \cap S^I} g^I(x) dF^I(x) + \int_{S^I \cap \overline{S^F}} g^I(x) dF^I(x) \right] \quad (13)$$

After some algebraic manipulations of Equation (13), the wage gap can simply be expressed as:¹⁰

$$\Delta = (\Delta_x + \Delta_F + \Delta_I) + \Delta_O \quad (14)$$

where Δ_x is the part of the wage gap that is explained by differences in characteristics of formal and informal workers over common support. Δ_F is the part of the wage gap explained by differences between two groups of formal workers i.e. those who have characteristics that can be matched to informal workers and those who cannot. Δ_I is the part of the wage gap explained by differences between two groups of informal workers: those who have characteristics that can be matched to formal workers and those who cannot. Δ_O accounts for the unexplained part of the wage gap. It is the part of the wage gap that cannot be explained by differences in characteristics of individuals and attributed to unobserved characteristics. The first three components on the right-hand side can be attributed to the wage gap that can be explained by differences in characteristics (explained wage gap) and the fourth term is attributed to the wage gap due to unobservable characteristics (unexplained wage gap) (Nopo, 2008).

6. RESULTS

6.1 The Determinants of Wages

We estimate the determinants of wages for 2004 and 2009 under the lower-bound and the upper-bound informality. The OLS regression estimates are presented in Appendix Table 4 (unadjusted for sample selection bias) and Table 5 (adjusted for sample selection). The top

¹⁰See Nopo (2008) for detail description of the method.

panel, (A), portrays the results for the year 2004 whilst panel (B) shows results for the 2009 data. The results show that, on average, the most important determinants of wages are education and experience. The coefficients of education dummies, as expected, are positive and significant in the formal sector in both samples and under both definitions of informality. The results show that formal workers tend to earn more as their education level increases. It is possible, therefore, to argue that well-educated formal workers tend to be more productive and earn more than workers with poorer education. Regarding the informal sector, the impact of education is found not to be strong as in the formal sector under the lower-bound informality. Contrary to expectations, the notable finding is that the returns to education are found to be high in the informal sector under the upper-bound definition of informality and the earning premium increases as the level of education increases. Returns to experience are hump-shaped, which show that experience significantly contributes earning at a certain point and then tends to decrease wages. Compared with the formal workers, for example, informal workers get, on average, an increase of over 4% for additional years of experience (in both samples) whilst formal workers get over 3% increase on their hourly wages under the upper-bound informality.

Next, we turn to the quantile regression results, which are presented in an Appendix (Tables 6 and 7). The results show that, at each quantile, coefficients of the education dummies, as expected, are positive and significant in the formal sector in 2004 as well as the 2009 sample. However, education affects wages differently in different parts of the wage distribution in the formal sector. Indeed, education tends to have a larger effect at higher quantiles. As their education level increases workers tend to earn more. In other words, we observe a trend suggesting a significant increase in the wage premium for highly educated workers. Hence, it could be hypothesised that there is a heterogeneity in the returns to educations, which tends to

be larger for formal workers at higher quantiles of the wage distribution (i.e. with better unobserved earning capacity). With respect to informal workers, some interesting results emerge. The coefficients of the education dummies are positive and significant in particular under the upper-bound informality. As before, the estimates show that the earning premium increases as the level of education increases. This could be an interesting result as it implies that for some informal workers education is an important tool for an income improvement. For example, compared with those holding primary education, workers who have a high school, vocational high school and university education tend to earn more at each quantile of the earning distribution under the upper-bound informality definition. This finding is observed when considering both the 2004 and 2009 data. By contrast, none of the coefficients of the education dummies are found to be significant in the informal sector in 2009 under the lower-bound definition. Education level, therefore, plays a very significant role in determining wages under the upper-bound informality in the informal sector, this is particularly true for the 50th, 75th and 90th quantiles. Experience contributes to better wages across the wage distribution in both the formal and the informal sector. The wage premium for being more experienced, however, is higher in the informal sector. For example, the wage premium for being more experienced is 4.4% at 90th quantile in the informal sector while 3.4% in the formal sector under the lower-bound informality in 2009. Similar results are obtained for the upper-bound informality and for the 2004 sample.

Other results are not surprising in the context of Turkey. The coefficient of male dummy varies across the wage distribution. It is positive and significant suggesting men enjoy a higher earning advantage over women both in the formal and the informal sector. The gender effect on the wage tends to become smaller at higher quantiles using both definitions of informality. However, the coefficient of the male dummy is much larger (in magnitude) in the

informal sector indicating women face more wage penalty in the informal sector. The results mainly show that married workers tend to earn more than singles in both formal and informal sectors. Looking at the employment location, we can say that, in general, both formal and informal workers in urban areas tend to earn more than workers in rural areas. With respect to the industrial sector, we find that workers in manufacturing and energy, and agriculture and mining sectors tend to earn less than workers in the finance and service sector both in the formal and the informal sector.

Overall, these results indicate that returns to education are high in the informal sector. There are several possible explanations for this result. First, this may be due to the fact that workers with high levels of education cannot find jobs that are suitable for them in the formal sector. Second, this might be related to the fact that small firms (indeed, informality is widespread among small firms in Turkey) pay a compensating differential to attract educated workers to compete with large firms. Firms in the formal sector must pay pension benefits, severance, frictions, social security contributions and so on. It seems possible that there may be a substitution, namely, firms that do not pay social security contributions may need to compensate well-educated workers more highly in order to make up for not giving them a benefit. Allowing a broader definition of the informal sector (upper-bound), there are formal workers who are moved into the informal sector and thereby resulting to the formal sector shrinking. As our results show, the returns to education are high in the informal sector under upper-bound informality. As such, having to rely on formal sector jobs could constitute an important loss of human capital when there is a movement between the formal sector and the informal sector. In this case, informal sector jobs could constitute an important potential of human capital.

Moreover, when the selection bias is accounted for, no substantial differences are observed for the explanatory variables and their effects on earnings. When contrasting the results of the wage equation estimates at the mean level it is clear that similar conclusion emerges irrespective of whether we adjust for sample selection or not. In other words, the strong selection of education at all levels into the formal sector remains consistent with higher returns to education in the formal sector and the informal sector under the upper-bound informality. The returns for university and vocational education are only slightly lower both in the formal and the informal sector. The returns to experience, on the other hand, are slightly higher when sample selection is taken into account. There is no discernible change observed in the informal sector. Regarding the gender variable, there are no significant differences between adjusted and unadjusted samples in 2004 and 2009 and, in both sectors. Men earn more than women in the formal sector under the lower-bound informality and in the informal sector under the upper-bound informality. In short, sample selection bias might not be an issue for concern after all in the context of Turkey.

6.2 The Size of the Wage Gap: The Oaxaca-Ransom Decomposition, the Machado and Mata (2005) and Nopo (2008) Decompositions

The results of the decomposition analysis based on the Oaxaca-Ransom technique are reported in Tables 3 through to 6. More specifically, Tables 3 and 4 present, respectively, the results for 2004 and 2009 without sample selection; whilst Tables 5 and 6 show the same results when sample selection is taken into account. The decomposition analysis shows that for the year 2004 the formal-informal wage gap is lower (0.675 log points) under the lower bound compared to the upper bound definition of informality (0.844). What is more, the results show that much of the wage gap (60% for the lower bound and 64% for the upper

bound) is explained by differences in characteristics of workers. Similar findings are also obtained with the 2009 data. It is worth noting, however, that the wage gap is slightly higher in 2009 than in 2004. As we will argue later, this difference may be due to the global financial crisis that started in 2007 or to changes in the size, composition and socio-economic structure of the population. We find that (after controlling for selection) differences in workers characteristics (the endowment effects) remain positive and statistically significant indicating that workers in the formal sector still have a higher earning advantage based on their observed characteristics. The notable difference is that the decomposition results show that more than half of the wage gap (54.5% in 2004 and 54% in 2009) is attributable to unobserved characteristics under the lower-bound definition of informality. In contrast to the lower-bound informality, differences in workers characteristics overwhelmingly explain the formal-informal wage differential under the upper-bound informality. This is similar to what has been found before controlling for selection.

“Place Table 3 about here”

“Place Table 4 about here”

“Place Table 5 about here”

“Place Table 6 about here”

Table 7 and Table 8 show results of the decomposition analysis using the MM technique for 2004 and 2009, respectively. In the first column of each table, we present the raw wage gap estimates-calculated as the difference in log hourly wages between formal and informal sector employees at specific quantiles of the wage distribution. In subsequent columns, estimates of the wage gap due to differences in observable characteristics (the covariates) and the differences in returns to those characteristics (the coefficients) are reported. The results are

also portrayed graphically; Figures 1-4 plot the estimated coefficient and covariate effects along the wage distribution. We find that for the two years the raw difference is positive across the entire wage distribution indicating that the wage difference in favour of workers in the formal sector. The results reveal that the formal-informal wage gap is sizeable especially at the bottom of the wage distribution for both years. The largest fraction of the formal-informal wage gap is attributable to both differences in workers characteristics and in the return to those characteristics. The unexplained component of the wage gap (coefficient effect) is larger at the lowest quantile, thus suggesting that the wage penalty faced by the informal worker is higher at the lowest quantile. The other notable finding is that the wage gap is higher under the upper-bound definition of informality. This difference could be explained by the fact that the lower-bound definition of informality gives us a minimum number of informal workers. Under this definition, more workers will find themselves in the formal sector and, thus, few workers will be in the informal sector. Hence, there are formal workers who could be classified as informal under the upper-bound definition of informality. When many low-wage formal workers are transferred from lower-bound to the upper-bound as informal, the average wage in the formal sector under the upper-bound informality rises, which may explain the large wage gap under the upper-bound informality.

Another important finding is that the size of the wage gap is higher in 2009. This difference may be due to changes in the two samples (i.e. socio-economic structure); it may also be explained by the global financial crisis started in 2007, which led to a decline in economic activities and job losses. Along this line, Yazir (2010) reports that the global financial crisis has led to an increase in unemployment in Turkey by 0.7% in 2008 and 3% in 2009. More workers, therefore, entered the informal sector. The change in the formal-informal wage gap between 2004 and 2009 may also be attributable to a high level of urbanisation in Turkey due

to a significant migration from rural to urban areas. Adaman and Kaya (2012) point out that Turkey has been experiencing large rural-urban migration and thereby rapid urbanisation since the 1950s. They state that urbanisation increased from 25% in 1950 to 44% in 1980 and 80% in 2010 and internal migrants are mainly involved in the informal sector. In this case, one would expect that the informal wage would decrease. Moreover, increased unemployment in Turkey (10.83% in 2004 and 14.03% in 2009) pushed workers towards the informal sector, thus resulting in the fall of the informal sector wage.

“Place Figure 1 about here”

“Place Figure 2 about here”

“Place Table 7 about here”

“Place Figure 3 about here”

“Place Figure 4 about here”

“Place Table 8 about here”

Finally taking advantage of the matching approach, which allows us to compare workers with similar characteristics in both sectors, we consider differences in the distribution of their observed characteristics. To this end, as aforementioned, we follow the non-parametric matching approach proposed by Nopo (2008). Table 9 presents the estimated results of a matching-based decomposition where matching is based on experience, experience squared, marital status, education dummies, industrial sector dummies and employment location. As explained previously, the total wage gap is divided into four additive components and expressed as: $\Delta = (\Delta F + \Delta X + \Delta I) + \Delta O$. The first three components on the right-hand side are attributed to differences in observable characteristics and the last component is attributed to differences in unobservable characteristics.

The matching results also indicate that Turkey stands out as a country with a positive formal-informal wage differential i.e., on average, formal workers earn more than informal workers although the gap is small. The total wage gap is found to be around 0.048 for the lower-bound informality and 0.060 for the upper-bound informality in 2004. The results, also, reveal contrasting differences for both samples. In fact, the results based on the 2004 sample, suggest that differences in observable characteristics explain much of the wage gap between the formal and informal sector. By contrast, results derived from the 2009 data indicate that a significant part of the wage gap comes from unobservables (ΔO)

Using the matching approach allows us to disentangle the wage differential. Hence it is interesting to probe further in some of the components, particularly ΔF and ΔI . The sign of the coefficient of ΔF (part of the wage difference explained by differences between formal workers in and out of the common support) is positive, which indicates that there are formal workers who have some characteristics that informal workers do not have and, these characteristics are highly rewarded in the Turkish labour market. In other words, these results demonstrate that unmatched formal workers earn more than matched formal workers on average. Furthermore, results from the 2009 sample reveal that an important part of the wage gap (0.395), which is explained, comes from differences between formal workers in and out of the common support. By contrast, a positive sign of the estimate of ΔI (part of the wage gap is attributed to differences in characteristics of informal workers in and out of common support) shows that matched informal workers earn more than unmatched informal workers.

Finally, the wage gap between the two sectors is also examined at different quantiles of the wage distribution after the matching procedure. This analysis is carried out by taking the difference between the wage distributions for the respective quantiles after matching. Table

10 reports the results. In line with the MM decomposition results, Nopo's matching decomposition also shows that informal workers face a wage penalty across the wage distribution.

“Place Table 9 about here”

Also, similar to the MM decomposition results, the wage gap changes across the distribution of wages and it is higher in the lowest quantiles of the earning distribution. These results confirm that the formal-informal wage gap in Turkey is driven by differences in earnings in the lower quantiles of the earning distribution and more visible under the upper-bound informality. However, when applying the matching method, it is important to note that there is a declining trend in the earning differences in the common support after matching. For example, in relation to the 2004 sample differences in earnings in support change from a coefficient of 0.801 and 0.916 to 0.466 and 0.549 at the 10th quantile for the lower-bound and the upper-bound informality. In the same way, there is a decline in earning differences after matching from 0.981 to a 0.841 under the lower-bound informality and from 0.981 to a 0.865 under the upper-bound informality at the 10th quantile in 2009. Comparisons between the decomposition results based on the quantile regression and the matching method suggest that, in general, there are no substantial differences. However, looking at the wage gap at different quantiles of the earning distribution in and out of the common support, it can be easily noticed that the wage gap is substantially higher out of the common support at all quantiles of the wage distribution. The formal-informal wage gap tends to be higher in the lowest quantiles and lower in the top quantiles of the earning distribution indicating higher informal wage penalty in the lowest quantiles.

“Place Table 10 about here”

7. CONCLUSIONS AND IMPLICATIONS

7.1 Summary of Key Findings and Implications

In this paper, we explored whether there is a wage gap between the formal and the informal sector in Turkey. For this purpose, a comprehensive empirical approach is used to derive the results of this paper. Indeed, as an initial step, the wage gap was investigated at the mean using the OLS regression and the Oaxaca-Ransom (1994) decomposition method-with and without the sample selection. The results derived from this technique indicate a wage gap in favour of formal sector workers and that the wage gap was higher under the upper bound definition of informality. Although the mean-based decomposition is a straightforward way to decompose the differences in wages it does not provide a full picture of the wage gap as it only explains mean differences in wages. In order to overcome the limitations of the mean decomposition, the quantile regression and the MM decomposition were applied to take account of heterogeneity in the wage distribution. The results reveal a wage gap is in favour of formal sector workers along the wage distribution. Moreover, we found that the wage gap was greater at the bottom of the wage distribution in both 2004 and 2009 samples. To ensure robustness of our findings we adopted a non-parametric method based on matching proposed by Nopo (2008), which compares comparable individuals effectively. This method measures the formal-informal wage gap only on the observations where characteristics of formal and informal workers are comparable (i.e. within the common support) and thereby addresses the problem of common support. The findings of the matching decomposition technique confirmed the existence of an informal sector wage penalty. The wage gap was higher especially in the lowest quantiles of the earning distribution in and out of the common

support. Nopo's decomposition method may be useful in helping to design more effective policies. The method allows formal workers to be matched to informal workers and thereby reducing differences between unmatched and matched workers. However, it might be useful to note that if there are two observationally equivalent workers, one working formally and the other informally, this may be the case-because formal jobs are "rationed"; but it may also be that there are unobservable, individual-specific characteristics such as ability or preferences for certain working conditions.¹¹

Taken together, the findings of this paper suggest that human capital factors (education and experience) have a key role in improving the wages. Most of the wage gap between the two sectors is explained by characteristics of workers and common characteristics that contribute to the wage gap are education and experience. The higher educated and more experienced workers are likely to have higher earnings both in the formal and the informal sector. Policies that promote education and equal opportunities for workers in the informal sector would improve earnings for many workers in Turkey by increasing productivity and incomes. Therefore, unlike Simsek (2014) claims, the Turkish government must attune their understanding of the role of the informal sector in Turkey. While efforts should be made to address negative aspects of the informal sector and support people in the informal sector, we argue that it is important not to destroy its capacity to provide livelihoods and develop output, employment and entrepreneurial potential. The potential of the informal sector cannot be overlooked.

¹¹ We are grateful to an anonymous referee for pointing this out.

7.2 Limitations and Directions for Further Research

In this paper, the analysis of wage gap between the two sectors is limited to examining wage earners based on cross-sectional data. Our results must be interpreted conditional on the wage earners. Alternatively, to capture variations in the wage differential a broader dataset, such as panel data, might be useful. Given the fact that the self-employed and unpaid family workers account for a large part of the informal sector in Turkey, an extension of the paper could possibly look to approach the issue from this angle.

REFERENCES

- Adaman, F. and Kaya, A. (2012) 'Social Impact of Emigration and Rural-Urban Migration in Central and Eastern Europe', Turkey Country Report, European Commission DG Employment and Social Affairs Inclusion.
- Altonji, J. and Blank, R. (1999) 'Race and Gender in the Labour Market' in O. Ashenfelter and D. Card (eds.) *Handbook of Labour Economics*. North Holland.
- Atiyas, A., Bakis, B. and Orhan, K.Y. (2017) 'Firms Dynamic and Job Creation in Turkey – Some Preliminary Results', REF Working Paper No.1.
- Aydin, E., Hisarciklilar, M. and Ilkcaracan, I. (2010) 'Formal versus Informal Labour Market Segmentation in Turkey in the Course of Market Liberalisation', *Topics in the Middle Eastern and North African Economies Electronic Journal*, 12:1-43.
- Badaoui, E., Strobl, E. and Walsh, F. (2008) 'Is There an Informal Employment Wage Penalty? Evidence from South Africa', *Economic Development and Cultural Change*, 56(3): 683-710.
- Bargain, O. and Kwenda, P. (2009), 'The Informal Sector Wage Gap: New Evidence Using Quantile Estimations on panel data', IZA Discussion Paper No. 4286.
- Bargain, O. and Kwenda, P. (2013) 'The Informal Sector: New Evidence Using Quantile Estimation on Panel Data', HAL Working Paper No.60.
- Barsky, R., Bound, J., Charles, K. K. and Lupton, J. P. (2002) 'Accounting for the Black-White Wealth Gap: A Non-Parametric Approach', *Journal of American Statistical Association*, 97(459):663-673.
- Baskaya, Y. S. & Hulagu, T. (2011). Informal-Formal Worker Wage Gap in Turkey: Evidence from A Semi-Parametric Approach, 51st Congress of the European Regional Science Association: New Challenges for European Regions and Urban Areas in a Globalised World, 30 August - 3 September 2011, Barcelona, Spain, European Regional Science Association (ERSA), Louvain-la-Neuve. Link: https://www.econstor.eu/dspace/bitstream/10419/120000/1/ERSA2011_0409.pdf (Accessed on January 20, 2018).
- Baslevant, C. and Acar, A. (2015) 'Recent Trends in Informal Employment in Turkey', *Yildiz Social Sciences Review*, 1(1):77-88.
- Ben Salem M. and Bensidoun I. (2012) 'The heterogeneity of informal employment and segmentation in the Turkish labour market', *Journal of the Asia Pacific Economy*, 17:578–592.
- Black, A. D., Haviland, M. A., Sanders, G. S. and Taylor, J. L. (2008) 'Gender Wage Disparities among the Highly Educated', *The Journal of Human Resources*, 43(3):630-659.
- Buchinsky, M. (1998) 'The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach', *Journal of Applied Econometrics*, 68(2): 303-338.

- Buehn, A. and Schneider, F. (2012) 'Shadow Economies Around the World: Novel Insights, Accepted Knowledge, and New Estimates', *International Tax and Public Finance*, 19:139-171.
- Buehn, A. and Schneider, F. (2013) 'Estimating the Size of the Shadow Economy: Methods, Problems and Open Questions', Johannes Kepler University of Linz Working Paper No.1320.
- Carneiro, G. C. and Henley A. (2001) 'Modelling Formal vs. Informal Employment and Earnings: Micro-Econometric Evidence for Brazil', The University of Wales, School of Business and Economics Research Paper No: 15.
- Cudeville, E. and Gurbuzer, Y. Z. (2010) 'Gender Wage Discrimination in the Turkish Labour Market: Can Turkey be Part of Europe?', *Comparative Economic Studies*, 52(3):429-463.
- Daza, N. and Gamboa, L. F. (2013) 'Informal-Formal Wage Gap in Colombia', ECINE Working Paper No.301.
- Ercan, H. (2011) 'Occupational Outlook in Turkey', Ankara: ILO Office for Turkey.
- Eurofound (2013) 'Tackling Undeclared Work in Turkey', Dublin.
- Fethi, M. D., Fethi, S., and Katircioglu, S. (2006), Estimating the Size of the Cypriot Underground Economy: A Comparison with European Experience, *International Journal of Manpower*, 27 (6): 515 - 534.
- Fethi, M. D., Fethi, S., & Katircioglu, S. (2004), Estimating the Underground Economy and Tax Evasion: Cointegration and Causality Evidence in the Case of Cyprus, 1960-2003, *International Journal of Economic Development*, 6 (3): 120-157.
- Fields, G. (1990) 'Labour Market Modelling and the Urban Informal Sector: Theory and Evidence' in D. Turnham, B. Salome and A. Schwartz (eds.) *The Informal Sector Revisited*. Paris: OECD
- Fortin, N., Lemieux, T. and Firpo, S. (2010), 'Decomposition Methods in Economics', NBER Working Paper, No. 16045.
- Goraus, K. (2013) 'Gender Gap in Labour Market Participation in Romania', *Timisoara Journal of Economic and Business*, 6(19):55-69.
- Gursel, S. (2014) 'Informal Employment in Turkish Economy,' Retrieved from <https://www.todayszaman.com>.
- Harries, J. and Todaro, P. M. (1970) 'Migration, Unemployment and Development: A Two-Sector Analysis', *The American Economic Review*, 60(1): 126-142.
- Heckman, J. J. (1979) 'Sample Selection Bias as a Specification Error', *Econometrica*, 47(2):153-161.

Henley, A, Arabsheibani, G. R. and Carneiro, F. G. 2006. 'On Defining and Measuring the Informal Sector', World Bank Policy Research Working Paper No.3866.

Henley A., Arabsheibani G. R. and Carneiro G. F. (2009) 'On Defining and Measuring the Informal Sector: Evidence from Brazil', *World Development*, 37(5):992-1003.

Huber, M. and Melly, B. (2011) 'Quantile Regression in the Presence of Sample Selection', University of St. Gallen Discussion Paper No.2011-09.

Hyder, A. and Reilly, B. (2005) 'The Public Sector Pay Gap in Pakistan: A Quantile Regression Analysis', PRUS Working Paper No.33.

ILO (2003) 'Labour Act of Turkey', Retrieved from <http://www.ilo.org/public/english/region/eurpro/ankara/legislation/law4857.htm>.

ILO (2012) 'The Youth Unemployment Crises: Time for Action', International Labour Office: Geneva.

Imamoglu, H., Katircioglu, S., & Payaslioglu, C.E. (2018), Financial Services Spillover Effects on Informal Economic Activity, *The Services Industries Journal* (in Press, <https://doi.org/10.1080/02642069.2017.1423056>).

Imamoglu, H. (2016), Re-estimation of the Size of Underground Economy in European Countries: MIMIC Approach, *International Journal of Economic Perspectives*, Volume 10, Issue 1, pp. 171-193.

Kan, E. O. and Tansel, A. (2014) 'Defining and Measuring Informality in the Turkish Labour Market,' MPRA Paper, No. 57739.

Lee, L. F. (1978) 'Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables', *International Economic Review*, 19(2):415-433.

Machado, S. and Mata, S. (2005) 'Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression', *Journal of Applied Econometrics*, 20(4):445-465.

Maloney, W. F. (1999) 'Does Informality Imply Segmentation in Urban Labour Markets? Evidence from Sectoral Transitions in Mexico', *World Bank Economic Review*, 13(2): 275-302.

Maloney, W. F. (2004) 'Informality Revisited', *World Development*, 32(7):1159-1178

Marcouiller, D., Castilla, R. V. and Woodruff, C. (1997) 'Formal Measures of the Informal Sector Wage Gap in Mexico, El Salvador and Peru', *Economic Development and Cultural Change*, 45(2): 367-392.

Mincer, J. (1974) 'Schooling, Experience and Earnings', New York: Colombia University Press.

Ministry of Development of Turkey (2012) 'Turkey's Sustainable Development: Claiming the Future', Ankara.

Mizala, A., Romaguera, P. and Gallegos, S. (2011) 'Public-Private Wage Gap in Latin America (1992-2007): A Matching Approach', *Labour Economics*, 18:115-131.

Monsted, T. (2000) 'Wage Differentials Between the Formal and the Informal Sector in Urban Bolivia', ILSEC-UCB Working Paper No.01/00

Montenegro, C. (2001) 'Wage Distribution in Chile: Does Gender Matter? A Quantile Regression Approach', World Bank Policy Research Working Paper No.20.

Nopo, H. (2008) 'Matching as a tool to decompose wage gap', *The Review of Economics and Statistics*, 90(2):290-299.

Oaxaca, R. L. (1973) 'Male-Female Wage Differentials in Urban Labour Markets', *International Economic Review*, 14(3):693-709.

Oaxaca, R. L. and Ransom, M. R. (1994) 'On Discrimination and the Decomposition of Wage Differentials', *Journal of Econometrics*, 61:5-21.

Perry, E., Maloney, W., Arias, O., Fajnzylber, P., Mason, A. and Saavedra, J. (2007) 'Informality: Exit and Exclusion', World Bank, Washington DC.

Portes, A., Castells, M. and Benton, L. A. (1989) 'Studies in Advanced and Less Developed Countries', John Hopkins University Press.

Pratap, S. and Quintin, E. (2006) 'Are Labour Markets Segmented in Developing Countries? A Semi-Parametric Approach', *European Economic Review*, 50:1817-1841.

Schneider, F., Buehn, A. and Montenegro, E. C. (2010) 'Shadow Economies All over the World: New Estimates for 162 Countries from 1999 to 2007', World Bank Policy Research Working Paper No.5356.

Schneider, F. and Williams, C. C. (2013) 'The Shadow Economy', The Institute of Economic Affairs.

Simsek, M. (2014) 'Economic Shadows and Light,' Retrieved from <https://www.project-syndicate.org>.

Sookram, S. and Watson, P. (2008) 'The Informal Sector Poverty and Gender in the Caribbean: The Case of Trinidad & Tobago', *Journal of Eastern Caribbean Studies*, 33(4): 43-68.

Staneva, A. and Arabshaibani, G. (2014) 'Is There an Informal Employment Wage Premium? Evidence from Tajikistan', *IZA Journal of Labour and Development*, 3(1):1-24.

Tannuri, E. M. and Pianto, M. D. (2002) 'Informal Employment in Brazil-A Choice at the Top and Segmentation at the Bottom: A Quantile Regression Approach', University of Brasilia Department of Economic Working Paper No.236.

Tansel, A. (1999) 'Formal versus Informal Sector Choice of Wage Earners and Their Wages in Turkey', Economic Research Forum Working Paper No.9927.

Tansel, A. (2000) 'Wage Earners, Self-Employment and Gender in the Informal Sector in Turkey', World Bank Policy Research Working Paper No. 24.

Tansel, A. and Kan, O. E. (2012) 'The Formal/Informal Employment Earnings Gap: Evidence from Turkey', MPRA Working Paper No.38498.

Turkish Statistical Institute (2013) 'Press Release on the Household Labour Force Survey', Ankara.

Van Der Velde, L., Tyrowicz, J. and Goraus, K. (2013) 'What is the True Gender Wage Gap? A Comparative Analysis Using Data from Poland', University of Warsaw Working Paper No.28.

Williams, C. C. (2008), Cross-National Variations in Undeclared Work: Results from a Survey of 27 European Countries, *International Journal of Economic Perspectives*, Volume 2, Issue 2, pp. 46-63.

Williams, C. C. & Round, J. (2009), Out of The Margins: Re-Theorizing the Role Of The Informal Economy In Ukraine, *International Journal of Economic Perspectives*, Volume 3, Issue 1, pp. 45-58.

Williams, C. C. (2010), Evaluating Competing Theories of the Shadow Economy: Some Lessons from an English Locality, *International Journal of Economic Perspectives*, Volume 4, Issue 2, pp. 391-402.

Williams, C. C. (2011a), Blurring the formal/informal economy divide: beyond a dual economies approach, *International Journal of Economic Perspectives*, Volume 5, Issue 4, pp. 347-362.

Williams, C. C. and Nadin, S. (2012a) 'Tackling the Hidden Enterprise Culture: Government Policies to Support the Formalization of Informal Enterprises', *Entrepreneurship and Regional Development*, 24(9-10):895-915.

Williams, C. C. and Nadin, S. (2012b) 'Work beyond Employment: Representations of Informal Economic Activities', *Work, Employment and Society*, 26(2):1-10.

Williams, C. C. (2013) 'Tackling Europe's Informal Economy: A Critical Evaluation of the Neo-Liberal Deregulatory Perspective', *Journal of Contemporary European Research*, 9(2):261-278.

Williams, C. C. (2011) 'Explaining the Perspective of the Informal Economy in Central and Eastern Europe: Some Lessons from Moscow', *Journal of the Economy and its Applications*, 1(1):22-52.

World Bank, (2010) 'Turkey Country Economic Memorandum: Informality: Causes, Consequences and Policies', The World Bank Report No.48523

World Bank, (2012) 'Towards Gender Equality in Turkey: A Summary Assessment', World Bank, Washington DC.

Yazir, S. (2010) ‘Adverse Impact of Global Financial Crisis on Employment and Social Security in Turkey’, Council of Europe SSC SSR Report.

Table 1: Lower-bound and upper-bound informality by age

	HBS 2004		HBS 2009	
	Lower-bound	Upper-bound	Lower-bound	Upper-bound
Age				
15-24	105	690	148	919
25-34	188	787	205	919
35-46	258	843	249	899
47-56	128	375	165	528
57-64	26	67	62	183
Total	705	2762	829	3448

Table 2: Distribution of wage-earners by informality status

	HBS 2004		HBS 2009	
	Formal	Informal	Formal	Informal
Lower-Bound	5104	705	6498	829
Upper-Bound	3047	2762	3879	3448
Social security registration	3649	2160	4756	2571
Contract status	4726	1083	6146	1181
Small firm employment	3833	1976	4747	2580
Total number of wage-earners		5809		7327

Table 3: 2004, Wage gap decomposition results

Upper-bound informality		Lower-bound informality	
Wage gap (100%)	0.844*** (0.0187)	Wage gap (100%)	0.675*** (0.0328)
Explained gap (64%)	0.540*** (0.0153)	Explained gap (60%)	0.408*** (0.0214)
Unexplained gap (36%)	0.304*** (0.0142)	Unexplained gap (40%)	0.267*** (0.0272)

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: 2009, Wage gap decomposition results

Upper-bound informality		Lower-bound informality	
Wage gap (100%)	0.885*** (0.0175)	Wage gap (100%)	0.746*** (0.0321)
Explained gap (64%)	0.562*** (0.0144)	Explained gap (61%)	0.458*** (0.0191)
Unexplained gap (36.5%)	0.323*** (0.0133)	Unexplained gap (38.5%)	0.287*** (0.0263)

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.

Table 5: 2004, Wage gap decomposition results corrected for selection bias

Lower-bound informality		Upper-bound informality	
Wage gap (100%)	0.875*** (0.0351)	Wage gap (100%)	0.883*** (0.0196)
Explained gap (45.5%)	0.398*** (0.0200)	Explained gap (61%)	0.539*** (0.0150)
Unexplained gap(54.5%)	0.477*** (0.0307)	Unexplained gap (38.8%)	0.343*** (0.0155)

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: 2009, Wage gap decomposition results corrected for selection bias

Lower-bound informality		Upper-bound informality	
Wage gap (100%)	0.945*** (0.0356)	Wage gap (100%)	0.908*** (0.0188)
Explained gap (46%)	0.435*** (0.0185)	Explained gap(62%)	0.561*** (0.0142)
Unexplained gap (54%)	0.510*** (0.0307)	Unexplained gap (38%)	0.347*** (0.0150)

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Figure 1: 2004, Lower-bound formal-informal wage differentials

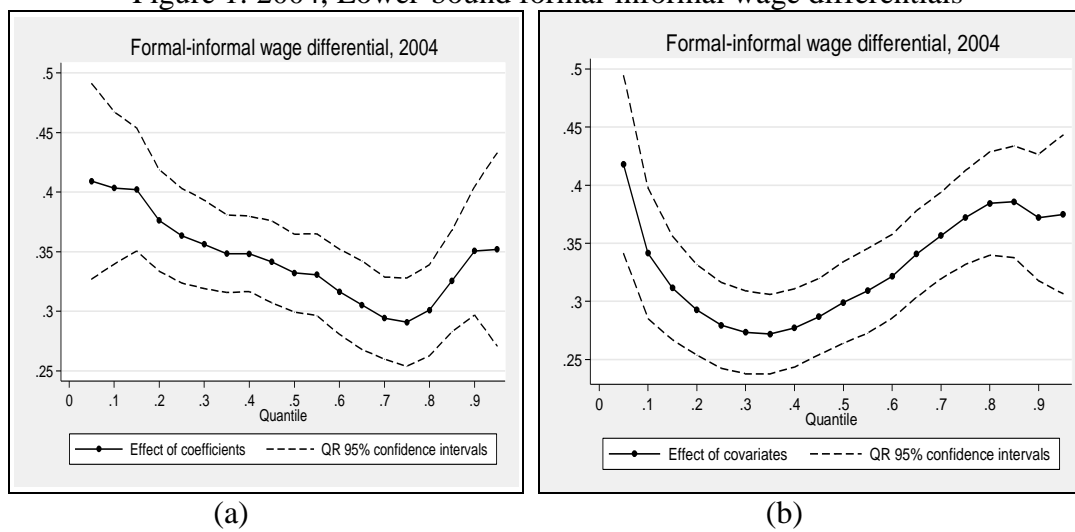


Figure 2: 2004, Upper-bound formal-informal wage differential

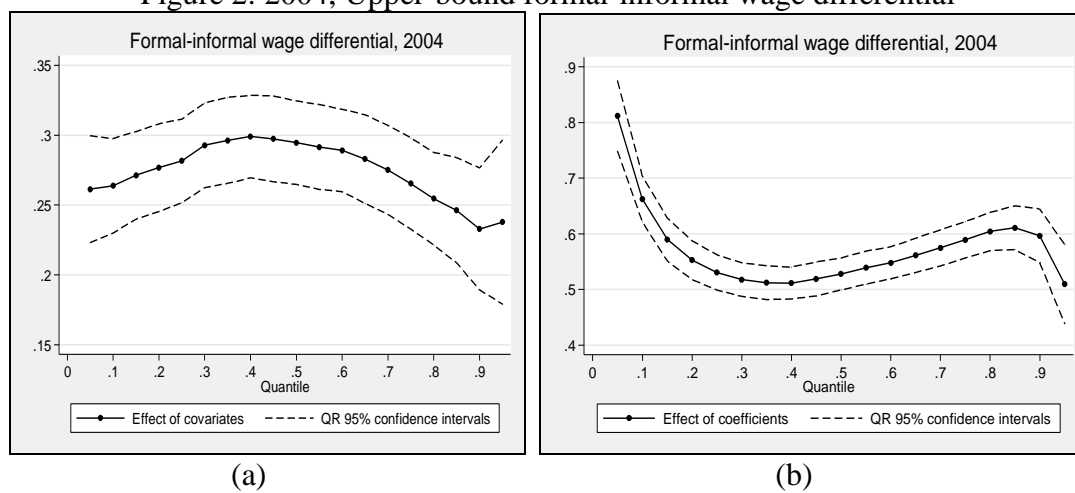


Table 7: 2004, MM formal-informal wage gap results

Panel A						
Lower-bound informality						
Quantiles	Raw gap	Coefficient effect	% Change	Covariate effect	% Change	Residuals
0.10	0.801	0.403 (0.033)	50.31%	0.342 (0.029)	42.69%	0.056
0.25	0.624	0.363 (0.020)	58.17%	0.279 (0.019)	44.71%	-0.019
0.50	0.554	0.332 (0.017)	60.00%	0.299 (0.018)	53.97%	-0.077
0.75	0.734	0.291 (0.019)	39.64%	0.372 (0.021)	50.68%	0.071
0.90	0.693	0.351 (0.028)	50.64%	0.372 (0.028)	53.68%	-0.030
Panel B						
Upper-bound informality						
Quantiles	Raw gap	Coefficient effect	% Change	Covariate effect	% Change	Residuals
0.10	0.916	0.662 (0.021)	72.27%	0.264 (0.017)	28.82%	-0.010
0.25	0.731	0.530 (0.016)	72.50%	0.282 (0.015)	38.57%	-0.081
0.50	0.854	0.528 (0.015)	61.83%	0.295 (0.015)	34.54%	0.031
0.75	0.892	0.589 (0.017)	66.03%	0.265 (0.017)	29.70%	0.038
0.90	0.801	0.596 (0.025)	74.41%	0.233 (0.022)	29.08%	-0.028

Note: Bootstrapped standard errors are given in parenthesis.

Figure 3: 2009, Lower-bound formal-informal wage differential

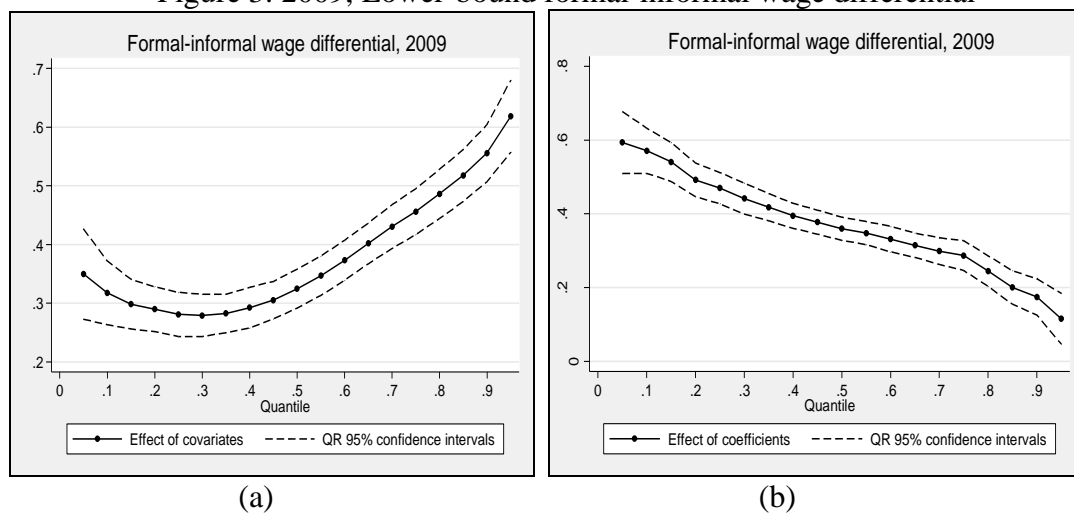


Figure 4: 2009, Upper-bound formal-informal wage differential

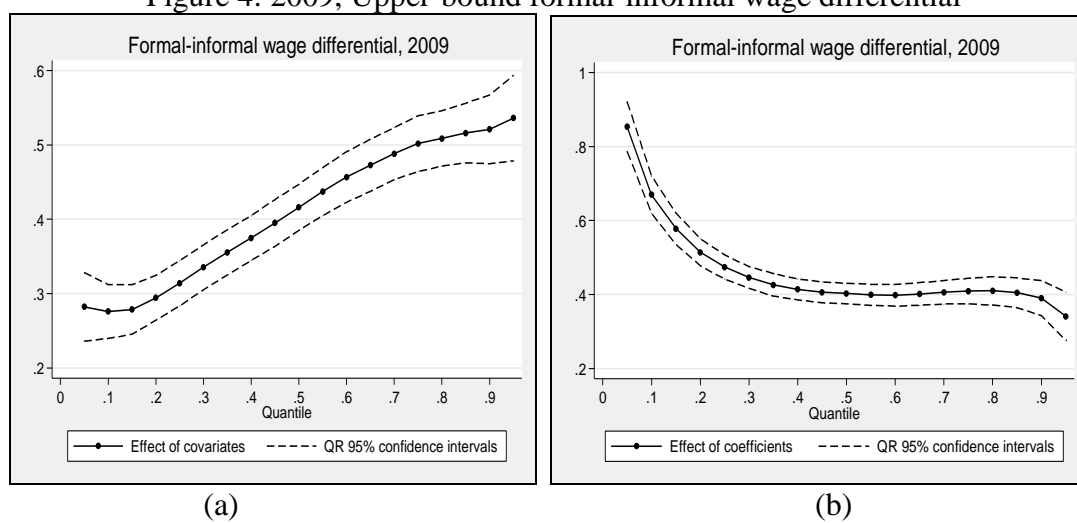


Table 8: 2009, MM formal-informal wage gap results

Panel A						
Lower-bound informality						
Quantiles	Raw gap	Coefficient effect	% Change	Covariate effect	% Change	Residual
0.1	0.981	0.571 (0.031)	58.20%	0.318 (0.028)	32.42%	0.092
0.25	0.742	0.470 (0.022)	63.34%	0.281 (0.019)	37.87%	-0.008
0.5	0.616	0.360 (0.016)	58.44%	0.325 (0.017)	52.76%	-0.068
0.75	0.815	0.287 (0.021)	34.21%	0.456 (0.020)	55.95%	0.072
0.9	0.778	0.175 (0.025)	22.49%	0.556 (0.025)	71.46%	0.048
Panel B						
Upper-bound informality						
Quantiles	Raw gap	Coefficient effect	% Change	Covariate effect	% Change	Residual
0.1	0.981	0.670 (0.026)	68.30%	0.276 (0.019)	28.13%	0.035
0.25	0.704	0.474 (0.017)	67.33%	0.314 (0.016)	44.60%	-0.084
0.5	0.799	0.403 (0.014)	50.44%	0.416 (0.016)	52.06%	-0.020
0.75	0.992	0.409 (0.018)	41.23%	0.502 (0.019)	50.60%	0.082
0.9	0.903	0.390 (0.024)	43.19%	0.521 (0.024)	57.70%	-0.009

Note: Bootstrapped standard errors are given in parenthesis

Table 9: 2004 and 2009 Nopo's matching based decomposition of the formal-informal wage gap

2004							
Lower-bound informality							
Δ	ΔO	ΔX	ΔF	ΔI	Matched workers	Matched formal	Matched informal
0.048	0.016	0.011	0.017	0.004	977	601 (61.5%)	376 (38.5%)
Upper-bound informality							
Δ	ΔO	ΔX	ΔF	ΔI	Matched workers	Matched formal	Matched informal
0.060	0.026	0.010	0.012	0.012	2606	1386 (53.2%)	1220 (46.8%)
2009							
Lower-bound informality							
Δ	ΔO	ΔX	ΔF	ΔI	Matched workers	Matched formal	Matched informal
1.260	0.547	0.278	0.395	0.041	2625	1971(75%)	654 (25%)
Upper-bound informality							
Δ	ΔO	ΔX	ΔF	ΔI	Matched workers	Matched formal	Matched informal
1.127	0.508	0.455	0.074	0.090	5561	3046 (54.8%)	2515 (45%)

Table 10: 2004 and 2009 Nopo's matching based decomposition of formal-informal wage gap by quantiles

On the common support				
2004		2009		
Percentiles	Lower-bound informality	Upper-bound informality	Lower-bound informality	Upper-bound informality
10	0.466	0.549	0.841	0.865
20	0.400	0.464	0.693	0.636
30	0.331	0.382	0.560	0.573
40	0.364	0.424	0.535	0.591
50	0.382	0.452	0.416	0.652
60	0.283	0.560	0.368	0.800
70	0.288	0.588	0.326	0.850
80	0.345	0.629	0.296	0.875
90	0.531	0.543	0.370	0.765
Out of the common support				
2004		2009		
Percentiles	Lower-bound informality	Upper-bound informality	Lower-bound informality	Upper-bound informality
10	1.022	1.211	1.381	1.480
20	0.901	1.061	1.086	1.139
30	0.717	1.135	0.827	1.086
40	0.671	1.150	0.758	1.235
50	0.731	1.179	0.811	1.335
60	0.845	1.150	0.885	1.322
70	0.912	1.140	1.002	1.291
80	0.841	1.099	0.981	1.316
90	0.721	0.984	0.808	1.141

APPENDIX

Table 1: Survey of studies on the formal-informal wage gap in Turkey

Author	Method	Definition of informality	Main findings
Aysit Tansel (1999)	OLS regression analysis with sample selection correction	Social security coverage	Results suggest significant informal wage penalty for both men and women
Aysit Tansel (2000)	OLS regression analysis with sample selection correction/ Oaxaca-Blinder decomposition method	Social security coverage	Results show that wages of men are higher in the formal sector than the informal sector and wages of self-employed individuals
Levent et al. (2004)	OLS regression analysis	Social security coverage and firm size	Significant informal wage penalty exists.
Hulagu and Baskaya (2011)	OLS regression analysis and Propensity Score Matching method (PSM)	Social security coverage	Results indicate significant informal wage penalty
Tansel and Kan (2012)	OLS regression analysis QR regression analysis	Social security coverage	Significant informal wage penalty was found using OLS. Results show significant informal wage penalty at lower quantiles

Table 2: Description of the variables

<i>Dependent Variables</i>	Definitions
Log wage	Log of the hourly wage obtained from individuals' monthly income from their main job.
<i>Independent Variables</i>	
Experience	age-schooling-6
<i>Sex</i>	
Male	dummy variable constructed as male=1 and female=0
<i>Education</i>	
No-education	=1 and zero otherwise if individuals are not able to read and write and individuals who are literate but have not completed any education.
Primary school	=1 and zero otherwise if individuals are completed 5 years primary school
Primary education	=1 and zero otherwise if individuals have completed 8 years primary education
High school	=1 and zero otherwise if individuals who have completed high school
Vocational high school	=1 and zero otherwise if individuals have completed any vocational or technical high school
University	=1 and zero otherwise if individuals completed any 2-year colleges, 4-year faculties and postgraduate (master and doctorate)
<i>Marital Status</i>	
Single	=1 and zero otherwise if individuals have never married
Married	=1 and zero otherwise if individuals are currently married to living together
Other marital status	=1 and zero otherwise if individuals are widowed and divorced and separate
<i>Employment Location</i>	
Urban	dummy variable constructed as urban=1 and rural=0
<i>Industry</i>	
Agriculture and mining	=1 and zero otherwise if individuals work in agriculture and mining sectors
Manufacturing and energy	=1 and zero otherwise if individuals work in manufacturing and energy sectors
Construction	=1 and zero otherwise if individuals work in the construction sector
Trade	=1 and zero otherwise if individuals work in the trade sector
Transportation	=1 and zero otherwise if individuals work in the transportation sector
Finance and other personals and social services	=1 and zero otherwise if individuals work in finance and service sectors.
<i>Other Variables</i>	
<i>Occupations</i>	
Professionals	=1 and zero otherwise if individuals are engaged in teaching professions, physical, mathematical and engineering professions and other professions
Legislators	=1 and zero otherwise if individuals are legislators, senior officials and managers

Associate professionals	=1 and zero otherwise if they are engaged in teaching, life science and other associate professions
Other professions	=1 and zero otherwise if individuals are engaged in skilled agricultural work, service work, craft and related works
Elementary profession	=1 and zero otherwise if individuals are engaged in occupations as drivers, labourers in mining, construction and others
<i>Measure of Informality</i>	
<i>Productive Definition</i>	
Contract	=1 and zero otherwise if individuals have a permanent job as their main job
Enterprise-based definition	
Small firm	=1 and zero otherwise if individuals are employed in small firms excluding groups of professionals and technicians who represent highly productive and high earning groups
<i>Legalistic Definitions</i>	
Social security registration	=1 and zero otherwise if individuals are registered for any social security registration
Estimates of Informality	
Upper-bound estimates of informality	=1 and zero otherwise if individuals are either not registered for any social security or do not have a contract or are employed in small firms
Lower-bound estimates of informality	=1 and zero otherwise if individuals are not covered by any social security and do not have a contract and are employed in small firms

Table 3: Descriptive statistics of the variables

Panel A- Lower-bound informality								
HBS 2004					HBS 2009			
	Formal		Informal		Formal		Informal	
Variables	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.
Log wage	14.61	0.79	13.93	0.82	1.34	0.83	0.59	0.87
Experience	19.65	10.88	25.89	12.26	19.07	11.39	25.34	12.94
Exp2	504.25	485.63	820.16	664.26	493.46	502.85	809.33	684.89
Male	0.79	0.41	0.70	0.46	0.75	0.43	0.60	0.49
Single	0.25	0.43	0.16	0.36	0.27	0.44	0.19	0.39
Married	0.73	0.44	0.80	0.40	0.70	0.46	0.73	0.44
Other m. status	0.02	0.14	0.04	0.20	0.03	0.17	0.08	0.27
No educ.	0.04	0.19	0.18	0.39	0.04	0.19	0.16	0.37
Primary sch.	0.34	0.47	0.62	0.48	0.29	0.45	0.60	0.49
Primary educ.	0.14	0.35	0.12	0.32	0.17	0.37	0.14	0.35
High sch.	0.22	0.42	0.06	0.23	0.14	0.35	0.05	0.21
Vocational sch.	0.90	0.28	0.02	0.13	0.13	0.34	0.04	0.19
University	0.18	0.38	0.00	0.05	0.23	0.42	0.01	0.08
Agricul/ mining	0.04	0.21	0.23	0.42	0.03	0.16	0.24	0.43
Manufacture/ enr	0.30	0.46	0.09	0.28	0.26	0.44	0.10	0.29
Construct	0.05	0.22	0.39	0.49	0.06	0.23	0.31	0.46
Trade	0.20	0.40	0.02	0.15	0.20	0.40	0.08	0.27
Transport	0.08	0.26	0.01	0.09	0.06	0.23	0.01	0.11
Finance and service	0.33	0.47	0.26	0.44	0.40	0.49	0.26	0.44
Urban	0.83	0.38	0.60	0.49	0.82	0.38	0.62	0.48
Total-N	5104		705		6498		829	
Panel B- Upper-bound informality								
HBS 2004					HBS 2009			
	Formal		Informal		Formal		Informal	
Variables	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.

Log wage	14.93	0.64	14.09	0.77	1.67	0.71	0.79	0.78
Experience	19.55	9.76	21.34	12.60	18.71	10.18	20.99	13.17
Exp2	477.56	417.71	614.34	607.02	453.70	429.05	614.15	624.35
Male	0.79	0.41	0.77	0.42	0.75	0.43	0.72	0.45
Single	0.19	0.39	0.29	0.45	0.22	0.41	0.30	0.46
Married	0.80	0.40	0.68	0.47	0.75	0.43	0.65	0.48
Other m. status	0.02	0.13	0.03	0.17	0.03	0.17	0.05	0.21
No educ.	0.01	0.11	0.10	0.30	0.01	0.11	0.10	0.30
Primary sch.	0.25	0.43	0.51	0.50	0.22	0.41	0.45	0.50
Primary educ.	0.11	0.31	0.17	0.37	0.12	0.32	0.21	0.41
High sch.	0.24	0.43	0.16	0.36	0.15	0.36	0.11	0.31
Vocational sch.	0.11	0.32	0.04	0.20	0.16	0.36	0.08	0.28
University	0.28	0.45	0.03	0.17	0.34	0.47	0.05	0.21
Agricul/ mining	0.02	0.13	0.12	0.33	0.02	0.12	0.09	0.29
Manufacture/ enr	0.33	0.47	0.20	0.40	0.27	0.45	0.20	0.40
Construct	0.03	0.16	0.17	0.37	0.03	0.17	0.15	0.36
Trade	0.12	0.33	0.25	0.43	0.14	0.35	0.24	0.43
Transport	0.06	0.24	0.07	0.26	0.03	0.18	0.07	0.26
Finance and service	0.44	0.50	0.19	0.39	0.50	0.50	0.24	0.43
urban	0.85	0.36	0.75	0.44	0.86	0.35	0.74	0.44
Total-N	3047		2762		3879		3448	

Table 4: OLS regression results (unadjusted for sample selection bias)

Panel A-2004				
	Lower-bound informality		Upper-bound informality	
	Formal workers	Informal workers	Formal workers	Informal workers
Experience	0.0519*** (0.00358)	0.0330** (0.0135)	0.0341*** (0.00427)	0.0415*** (0.00561)
Exp2	-0.001*** (7.11e-05)	-0.000** (0.000221)	-0.000*** (9.20e-05)	-0.000*** (9.88e-05)
Male	0.108*** (0.0224)	0.101 (0.0811)	0.0281 (0.0241)	0.223*** (0.0350)
Married	0.192*** (0.0290)	0.0947 (0.127)	0.181*** (0.0315)	0.125** (0.0488)
Other m. status	0.0947 (0.0646)	0.126 (0.194)	0.0366 (0.0768)	0.171* (0.0891)
Primary sch.	0.328*** (0.0487)	0.238*** (0.0853)	-0.00420 (0.0852)	0.303*** (0.0489)
Primary educ.	0.489*** (0.0528)	0.328*** (0.125)	0.145* (0.0882)	0.384*** (0.0596)
High sch.	0.847*** (0.0519)	0.0830 (0.153)	0.397*** (0.0869)	0.701*** (0.0613)
Vocational sch.	0.910*** (0.0566)	0.713*** (0.242)	0.509*** (0.0894)	0.612*** (0.0824)
University	1.400*** (0.0541)	0.713 (0.550)	0.893*** (0.0874)	1.267*** (0.0909)
Agricul/mining	-0.413*** (0.0458)	-0.201** (0.0953)	0.309*** (0.0709)	-0.329*** (0.0536)
Manufacture/enr	-0.105*** (0.0234)	-0.541*** (0.115)	-0.158*** (0.0234)	-0.0229 (0.0433)
Construct	-0.172*** (0.0417)	0.131 (0.0863)	0.0828 (0.0590)	-0.0283 (0.0476)

Trade	-0.368*** (0.0257)	-0.283 (0.206)	-0.309*** (0.0310)	-0.151*** (0.0419)
Transport	-0.0139 (0.0350)	-0.364 (0.325)	0.0633 (0.0396)	0.140** (0.0593)
Urban	0.0390* (0.0229)	0.134* (0.0698)	0.0104 (0.0253)	0.0809** (0.0342)
Constant	13.17*** (0.0623)	13.07*** (0.172)	13.93*** (0.0931)	12.96*** (0.0794)

Panel B-2009

	Lower-bound informality		Upper-bound informality	
	Formal workers	Informal workers	Formal workers	Informal workers
Experience	0.0501*** (0.00290)	0.0567*** (0.0110)	0.0345*** (0.00358)	0.0431*** (0.00453)
Exp2	-0.001*** (5.87e-05)	-0.001*** (0.000180)	-0.000*** (7.88e-05)	-0.001*** (8.04e-05)
Male	0.192*** (0.0191)	0.441*** (0.0802)	0.157*** (0.0212)	0.300*** (0.0299)
Married	0.188*** (0.0243)	-0.0205 (0.106)	0.154*** (0.0263)	0.0922** (0.0418)
Other m. status	0.120** (0.0493)	0.211 (0.146)	0.0642 (0.0567)	0.198*** (0.0698)
Primary sch.	0.247*** (0.0431)	0.0406 (0.0796)	0.149* (0.0794)	0.124*** (0.0445)
Primary educ.	0.374*** (0.0455)	0.0822 (0.111)	0.323*** (0.0813)	0.162*** (0.0514)
High sch.	0.714*** (0.0468)	0.264* (0.152)	0.542*** (0.0811)	0.499*** (0.0583)
Vocational sch.	0.727*** (0.0472)	0.329** (0.165)	0.567*** (0.0811)	0.455*** (0.0614)
University	1.371*** (0.0464)	0.648* (0.363)	1.150*** (0.0805)	0.805*** (0.0724)
Agricul/mining	-0.387*** (0.0499)	-0.488*** (0.0924)	-0.207*** (0.0718)	-0.357*** (0.0495)
Manufacture/enr	-0.217***	-0.958***	-0.278***	-0.189***

	(0.0213)	(0.104)	(0.0228)	(0.0374)
Construct	-0.223***	-0.228**	-0.246***	-0.0533
	(0.0354)	(0.0982)	(0.0517)	(0.0419)
Trade	-0.331***	-0.461***	-0.308***	-0.161***
	(0.0224)	(0.127)	(0.0270)	(0.0361)
Transport	-0.418***	-0.276	-0.406***	-0.223***
	(0.0359)	(0.246)	(0.0504)	(0.0517)
Urban	0.103***	0.0894	0.0555**	0.121***
	(0.0209)	(0.0663)	(0.0251)	(0.0296)
Constant	-0.0667	-0.193	0.424***	-0.157**
	(0.0526)	(0.156)	(0.0869)	(0.0671)
Observations	6,498	829	3,879	3,448

Note: Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 5: OLS regression results (adjusted for sample selection bias)

Panel A-2004				
Variables	Lower-bound informality		Upper-bound informality	
	Formal workers	Informal workers	Formal workers	Informal workers
Experience	0.041*** (0.004)	0.033** (0.014)	0.030*** (0.005)	0.041*** (0.006)
Exp2	-0.001*** (0.000)	-0.000** (0.000)	-0.000** (0.000)	-0.001*** (0.000)
Male	0.083*** (0.023)	0.104 (0.083)	0.011 (0.025)	0.229*** (0.037)
Married	0.139*** (0.028)	0.096 (0.126)	0.160*** (0.033)	0.129*** (0.046)
Other m. status	0.056 (0.064)	0.129 (0.211)	0.011 (0.072)	0.182* (0.094)
Primary sch.	0.288*** (0.058)	0.236*** (0.088)	-0.011 (0.107)	0.297*** (0.053)
Primary educ.	0.406*** (0.060)	0.325*** (0.121)	0.137 (0.107)	0.374*** (0.061)
High sch.	0.688*** (0.060)	0.076 (0.136)	0.370*** (0.106)	0.686*** (0.064)
Vocational sch.	0.725*** (0.063)	0.689* (0.400)	0.482*** (0.109)	0.588*** (0.085)
University	1.151*** (0.063)	0.712*** (0.147)	0.850*** (0.108)	1.232*** (0.105)
Urban	0.031 (0.023)	0.134* (0.069)	0.006 (0.025)	0.083** (0.036)
Agricul/mining	-0.306*** (0.062)	-0.197** (0.090)	0.325*** (0.097)	-0.317*** (0.058)
Manufacture/enr	-0.116*** (0.022)	-0.543*** (0.129)	-0.162*** (0.024)	-0.027 (0.045)
Construct	-0.040 (0.041)	0.135* (0.077)	0.095 (0.061)	-0.019 (0.047)
Trade	-0.259*** (0.026)	-0.281* (0.146)	-0.297*** (0.034)	-0.145*** (0.042)
Transport	0.065* (0.037)	-0.366 (0.343)	0.065 (0.044)	0.146** (0.062)
Lambda	0.417*** (0.022)	0.037 (0.194)	0.230*** (0.035)	0.091* (0.053)

Constant	13.624*** (0.080)	13.073*** (0.176)	14.044*** (0.118)	12.951*** (0.087)
Observations	5,104	705	3,047	2,762
R-squared	0.465	0.152	0.404	0.196
Panel B-2009				
Lower-bound informality		Upper-bound informality		
	Formal workers	Informal workers	Formal workers	Informal workers
Experience	0.040*** (0.003)	0.057*** (0.012)	0.032*** (0.004)	0.043*** (0.005)
Exp2	-0.001*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)
Male	0.163*** (0.019)	0.442*** (0.084)	0.147*** (0.021)	0.300*** (0.032)
Married	0.144*** (0.024)	-0.021 (0.117)	0.140*** (0.029)	0.093** (0.042)
Other m. status	0.088** (0.042)	0.212 (0.162)	0.048 (0.054)	0.199*** (0.066)
Primary sch.	0.204*** (0.048)	0.040 (0.077)	0.141* (0.079)	0.123*** (0.048)
Primary educ.	0.296*** (0.051)	0.081 (0.109)	0.312*** (0.081)	0.161*** (0.055)
High sch.	0.566*** (0.052)	0.262 (0.193)	0.520*** (0.082)	0.498*** (0.062)
Vocational sch.	0.560*** (0.052)	0.327* (0.181)	0.543*** (0.082)	0.453*** (0.061)
University	1.117*** (0.055)	0.644 (0.435)	1.113*** (0.083)	0.803*** (0.100)
Urban	0.072*** (0.020)	0.089 (0.068)	0.047** (0.023)	0.121*** (0.029)
Agricul/mining	-0.327*** (0.054)	-0.487*** (0.094)	-0.190*** (0.069)	-0.356*** (0.050)
Manufacture/enr	-0.227*** (0.023)	-0.958*** (0.129)	-0.279*** (0.024)	-0.189*** (0.043)
Construct	-0.112*** (0.033)	-0.227** (0.109)	-0.230*** (0.052)	-0.053 (0.042)
Trade	-0.257*** (0.023)	-0.460*** (0.138)	-0.302*** (0.029)	-0.160*** (0.036)
Transport	-0.336***	-0.274	-0.399***	-0.223***

	(0.037)	(0.227)	(0.050)	(0.051)
Lambda	0.405***	0.012	0.148***	0.006
	(0.020)	(0.141)	(0.033)	(0.042)
Constant	0.396***	-0.193	0.508***	-0.157**
	(0.065)	(0.161)	(0.092)	(0.075)
Observations	6,498	829	3,879	3,448
R-squared	0.477	0.214	0.449	0.173

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6:2004, Quantile regression estimates for formal and informal workers

Panel A										
Lower-bound informality										
Formal workers					Informal workers					
Variables	q10	q25	q50	q75	q90	q10	q25	q50	q75	q90
Experience	0.067*** (0.007)	0.062*** (0.005)	0.053*** (0.005)	0.040*** (0.005)	0.041*** (0.007)	0.022 (0.024)	0.018 (0.020)	0.033** (0.014)	0.040*** (0.015)	0.055*** (0.020)
Exp2	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.000** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.000** (0.000)	-0.001** (0.000)	-0.001** (0.000)
Male	0.126*** (0.041)	0.096*** (0.025)	0.078*** (0.023)	0.084*** (0.026)	0.054 (0.047)	0.394*** (0.123)	0.322*** (0.106)	0.148 (0.094)	-0.089 (0.101)	-0.293** (0.137)
Married	0.217*** (0.057)	0.171*** (0.033)	0.165*** (0.023)	0.213*** (0.036)	0.252*** (0.052)	0.519** (0.206)	0.108 (0.173)	0.065 (0.137)	0.031 (0.141)	-0.201 (0.265)
Other m. status	0.212 (0.129)	0.032 (0.067)	0.029 (0.060)	0.057 (0.074)	0.071 (0.127)	0.592 (0.424)	0.229 (0.228)	0.010 (0.222)	-0.066 (0.321)	-0.207 (0.505)
Primary sch.	0.314*** (0.102)	0.304*** (0.067)	0.316*** (0.052)	0.299*** (0.086)	0.234** (0.093)	0.235 (0.146)	0.318** (0.133)	0.210** (0.097)	0.211* (0.127)	0.211* (0.128)
Primary educ.	0.481*** (0.107)	0.412*** (0.071)	0.477*** (0.061)	0.493*** (0.090)	0.368*** (0.097)	0.419** (0.184)	0.353** (0.178)	0.370*** (0.132)	0.344** (0.167)	0.157 (0.172)
High sch.	0.791*** (0.108)	0.779*** (0.070)	0.813*** (0.056)	0.765*** (0.089)	0.701*** (0.101)	0.519** (0.231)	0.213 (0.178)	0.072 (0.138)	0.009 (0.211)	0.142 (0.208)
Vocational sch.	0.828*** (0.115)	0.834*** (0.077)	0.895*** (0.060)	0.959*** (0.096)	0.865*** (0.102)	0.573* (0.327)	0.553 (0.343)	0.370 (0.395)	0.442 (0.574)	0.932 (0.941)
University	1.387*** (0.106)	1.336*** (0.071)	1.327*** (0.056)	1.382*** (0.091)	1.385*** (0.107)	1.333** (0.535)	1.064*** (0.410)	0.811*** (0.287)	0.504** (0.221)	0.094 (0.176)
Agricul/mining	-0.847*** (0.122)	-0.607*** (0.089)	-0.461*** (0.075)	-0.184** (0.080)	-0.011 (0.131)	-0.462*** (0.163)	-0.413*** (0.122)	-0.033 (0.099)	-0.103 (0.141)	0.077 (0.160)
Manufacture/enr	-0.044 (0.039)	-0.100*** (0.028)	-0.128*** (0.025)	-0.097*** (0.027)	-0.066 (0.047)	-0.855*** (0.208)	-0.738*** (0.194)	-0.421** (0.211)	-0.422*** (0.154)	-0.380*** (0.119)
Construct	-0.308*** (0.067)	-0.237*** (0.054)	-0.165*** (0.046)	-0.091* (0.051)	0.010 (0.081)	-0.004 (0.126)	0.035 (0.093)	0.211*** (0.079)	0.085 (0.098)	0.254* (0.142)
Trade	-0.365*** (0.044)	-0.381*** (0.030)	-0.406*** (0.027)	-0.326*** (0.029)	-0.266*** (0.059)	-0.027 (0.250)	-0.307 (0.202)	-0.270 (0.179)	-0.392 (0.245)	-0.298 (0.323)
Transport	-0.132** (0.058)	-0.117*** (0.042)	-0.075* (0.043)	0.081* (0.045)	0.095 (0.062)	-0.859 (0.677)	-0.543 (0.649)	-0.358 (0.500)	-0.005 (0.390)	-0.149 (0.337)

Urban	0.094** (0.043)	0.043 (0.030)	0.021 (0.025)	0.026 (0.026)	-0.009 (0.048)	0.132 (0.141)	0.119 (0.081)	0.220*** (0.066)	0.154* (0.085)	0.197* (0.111)
Constant	12.392*** (0.118)	12.868*** (0.091)	13.263*** (0.083)	13.568*** (0.105)	13.891*** (0.118)	11.922*** (0.318)	12.651*** (0.245)	12.951*** (0.187)	13.623*** (0.212)	13.995*** (0.283)
Observations	5,104	5,104	5,104	5,104	5,104	705	705	705	705	705

Panel B										
Upper-bound informality										
	Formal workers					Informal workers				
	q10	q25	q50	q75	q90	q10	q25	q50	q75	q90
Experience	0.049*** (0.008)	0.037*** (0.006)	0.027*** (0.005)	0.018*** (0.006)	0.030*** (0.010)	0.045*** (0.009)	0.043*** (0.006)	0.038*** (0.006)	0.043*** (0.007)	0.050*** (0.009)
Exp2	-0.001*** (0.000)	-0.001*** (0.000)	-0.000** (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Male	0.075** (0.030)	0.083*** (0.027)	0.037 (0.025)	0.025 (0.034)	0.014 (0.063)	0.432*** (0.072)	0.304*** (0.053)	0.144*** (0.038)	0.080** (0.040)	0.136** (0.055)
Married	0.106*** (0.038)	0.135*** (0.036)	0.195*** (0.030)	0.300*** (0.041)	0.215*** (0.073)	0.191** (0.083)	0.115** (0.055)	0.129*** (0.046)	0.089* (0.051)	0.122* (0.065)
Other m. status	0.122 (0.090)	0.039 (0.083)	0.073 (0.077)	0.085 (0.106)	-0.049 (0.179)	0.180 (0.162)	0.182* (0.108)	0.083 (0.109)	0.088 (0.098)	0.161 (0.220)
Primary sch.	0.083 (0.133)	0.103 (0.120)	-0.027 (0.143)	-0.036 (0.122)	0.040 (0.226)	0.359*** (0.091)	0.355*** (0.072)	0.255*** (0.059)	0.329*** (0.075)	0.157* (0.083)
Primary educ.	0.252* (0.135)	0.287** (0.122)	0.173 (0.144)	0.111 (0.123)	0.148 (0.232)	0.468*** (0.109)	0.409*** (0.081)	0.313*** (0.067)	0.422*** (0.084)	0.291*** (0.096)
High sch.	0.483*** (0.135)	0.519*** (0.121)	0.374*** (0.143)	0.295** (0.122)	0.385* (0.232)	0.766*** (0.109)	0.649*** (0.080)	0.573*** (0.075)	0.698*** (0.088)	0.653*** (0.110)
Vocational sch.	0.588*** (0.140)	0.583*** (0.126)	0.494*** (0.145)	0.540*** (0.127)	0.581** (0.231)	0.612*** (0.156)	0.600*** (0.089)	0.454*** (0.088)	0.645*** (0.125)	0.519*** (0.154)
University	0.936*** (0.136)	0.956*** (0.122)	0.854*** (0.145)	0.871*** (0.125)	1.056*** (0.232)	1.245*** (0.159)	1.120*** (0.119)	1.117*** (0.103)	1.242*** (0.183)	1.682*** (0.269)
Agricul/mining	0.002 (0.164)	0.183 (0.120)	0.322*** (0.123)	0.540*** (0.116)	0.450** (0.192)	-0.546*** (0.107)	-0.351*** (0.082)	-0.203*** (0.062)	-0.224*** (0.075)	-0.335*** (0.120)
Manufacture/enr	-0.226*** (0.028)	-0.224*** (0.033)	-0.191*** (0.028)	-0.109*** (0.033)	-0.031 (0.045)	0.053 (0.090)	0.108** (0.049)	0.080* (0.047)	-0.084* (0.047)	-0.222** (0.104)
Construct	-0.301** (0.153)	-0.010 (0.108)	0.120* (0.072)	0.282*** (0.094)	0.142** (0.069)	-0.014 (0.099)	0.011 (0.048)	0.077 (0.052)	-0.030 (0.054)	-0.191* (0.107)

Trade	-0.391*** (0.043)	-0.381*** (0.041)	-0.366*** (0.038)	-0.240*** (0.053)	-0.067 (0.088)	-0.049 (0.079)	-0.048 (0.041)	-0.095** (0.048)	-0.196*** (0.049)	-0.382*** (0.097)
Transport	-0.147** (0.057)	-0.065 (0.068)	0.108** (0.043)	0.128** (0.050)	0.119 (0.096)	0.140 (0.096)	0.154*** (0.049)	0.111* (0.062)	0.068 (0.078)	0.128 (0.151)
Urban	0.024 (0.053)	0.015 (0.033)	0.007 (0.025)	0.016 (0.038)	-0.015 (0.061)	0.122** (0.062)	0.152*** (0.038)	0.117*** (0.036)	0.067 (0.045)	-0.003 (0.071)
Constant	13.278*** (0.163)	13.547*** (0.134)	13.985*** (0.149)	14.238*** (0.132)	14.353*** (0.239)	11.906*** (0.154)	12.412*** (0.099)	13.052*** (0.106)	13.423*** (0.108)	13.912*** (0.152)
Observations	3,047	3,047	3,047	3,047	3,047	2,762	2,762	2,762	2,762	2,762

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: 2009, Quantile regression estimates for formal and informal workers-lower-bound informality

Panel A										
Lower-bound informality										
Variables	Formal workers					Informal workers				
	q10	q25	q50	q75	q90	q10	q25	q50	q75	q90
Experience	0.064*** (0.005)	0.056*** (0.005)	0.047*** (0.003)	0.041*** (0.004)	0.034*** (0.006)	0.100*** (0.026)	0.056*** (0.016)	0.050*** (0.013)	0.054*** (0.016)	0.044* (0.023)
Exp2	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.000** (0.000)	-0.002*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001* (0.000)
Male	0.224*** (0.037)	0.159*** (0.028)	0.176*** (0.016)	0.194*** (0.023)	0.204*** (0.026)	0.500*** (0.148)	0.413*** (0.113)	0.384*** (0.084)	0.317*** (0.106)	0.369* (0.196)
Married	0.265*** (0.041)	0.197*** (0.038)	0.143*** (0.021)	0.167*** (0.023)	0.171*** (0.036)	-0.264 (0.225)	0.062 (0.134)	0.021 (0.131)	-0.105 (0.155)	-0.058 (0.207)
Other m. status	0.252*** (0.080)	0.155** (0.062)	0.040 (0.041)	0.047 (0.055)	0.070 (0.075)	-0.252 (0.380)	0.438** (0.197)	0.282 (0.184)	0.153 (0.191)	0.094 (0.252)
Primary sch.	0.261* (0.150)	0.191*** (0.062)	0.207*** (0.060)	0.153*** (0.059)	0.162** (0.074)	-0.106 (0.156)	-0.021 (0.106)	0.066 (0.086)	-0.003 (0.099)	0.043 (0.156)
Primary educ.	0.322** (0.158)	0.312*** (0.063)	0.355*** (0.062)	0.354*** (0.061)	0.330*** (0.075)	0.005 (0.252)	0.142 (0.157)	0.111 (0.127)	-0.007 (0.136)	0.040 (0.231)
High sch.	0.644*** (0.157)	0.638*** (0.068)	0.696*** (0.064)	0.614*** (0.063)	0.613*** (0.075)	-0.003 (0.235)	-0.116 (0.271)	0.254 (0.195)	0.126 (0.212)	0.215 (0.857)
Vocational sch.	0.714*** (0.158)	0.681*** (0.065)	0.673*** (0.065)	0.641*** (0.061)	0.671*** (0.077)	0.093 (0.411)	0.329 (0.260)	0.428*** (0.155)	0.158 (0.244)	0.556 (0.461)
University	1.302*** (0.158)	1.304*** (0.066)	1.321*** (0.063)	1.267*** (0.064)	1.400*** (0.078)	0.384 (0.641)	0.677 (0.702)	0.755 (0.593)	0.492 (0.609)	0.927 (0.645)
Agricul/mining	-0.584*** (0.124)	-0.450*** (0.086)	-0.366*** (0.057)	-0.386*** (0.054)	-0.229** (0.109)	-0.456*** (0.159)	-0.413*** (0.138)	-0.414*** (0.096)	-0.466*** (0.125)	-0.697*** (0.204)
Manufacture/enr	-0.150*** (0.036)	-0.127*** (0.029)	-0.227*** (0.023)	-0.273*** (0.026)	-0.223*** (0.030)	-1.373*** (0.263)	-1.057*** (0.286)	-0.876*** (0.136)	-0.828*** (0.187)	-0.705** (0.327)
Construct	-0.183*** (0.061)	-0.176*** (0.040)	-0.260*** (0.031)	-0.247*** (0.043)	-0.228*** (0.044)	-0.294 (0.203)	-0.089 (0.138)	-0.122 (0.109)	-0.101 (0.145)	-0.367* (0.215)
Trade	-0.262*** (0.035)	-0.299*** (0.029)	-0.380*** (0.024)	-0.389*** (0.029)	-0.303*** (0.045)	-0.664** (0.292)	-0.317** (0.154)	-0.288** (0.140)	-0.454*** (0.169)	-0.619** (0.263)
Transport	-0.416*** (0.072)	-0.402*** (0.057)	-0.410*** (0.035)	-0.442*** (0.041)	-0.341*** (0.062)	-0.095 (0.390)	-0.378 (0.310)	-0.322 (0.264)	-0.507 (0.389)	-0.894 (0.715)

Urban	0.128*** (0.034)	0.113*** (0.026)	0.109*** (0.022)	0.093*** (0.024)	0.077** (0.031)	0.099 (0.130)	0.125 (0.095)	0.079 (0.070)	0.128 (0.085)	0.151 (0.126)
Constant	-0.925*** (0.165)	-0.411*** (0.071)	0.073 (0.071)	0.491*** (0.070)	0.737*** (0.086)	-1.361*** (0.299)	-0.719*** (0.225)	-0.200 (0.182)	0.378* (0.193)	0.895*** (0.341)
Observations	6,498	6,498	6,498	6,498	6,498	829	829	829	829	829

Panel B										
Upper-bound informality										
Formal workers					Informal workers					
	q10	q25	q50	q75	q90	q10	q25	q50	q75	q90
Experience	0.050*** (0.007)	0.042*** (0.005)	0.029*** (0.004)	0.020*** (0.005)	0.027*** (0.006)	0.063*** (0.012)	0.046*** (0.006)	0.036*** (0.005)	0.040*** (0.005)	0.041*** (0.007)
Exp2	-0.001*** (0.000)	-0.001*** (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000 (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Male	0.134*** (0.033)	0.113*** (0.029)	0.151*** (0.020)	0.168*** (0.026)	0.172*** (0.032)	0.562*** (0.069)	0.371*** (0.046)	0.222*** (0.033)	0.217*** (0.033)	0.229*** (0.053)
Married	0.106** (0.051)	0.107*** (0.035)	0.172*** (0.026)	0.180*** (0.034)	0.161*** (0.041)	0.144 (0.108)	0.093* (0.055)	0.100** (0.043)	0.051 (0.043)	0.135** (0.060)
Other m. status	0.070 (0.101)	0.020 (0.060)	0.030 (0.050)	0.065 (0.079)	0.081 (0.092)	0.351** (0.162)	0.276*** (0.079)	0.143** (0.072)	0.167 (0.102)	0.110 (0.092)
Primary sch.	0.068 (0.177)	0.091 (0.083)	0.160 (0.112)	0.089 (0.090)	0.134 (0.086)	0.134 (0.153)	0.102 (0.066)	0.122*** (0.046)	0.085* (0.051)	0.080 (0.087)
Primary educ.	0.241 (0.181)	0.262*** (0.088)	0.381*** (0.113)	0.251*** (0.087)	0.223*** (0.082)	0.105 (0.158)	0.078 (0.077)	0.178*** (0.053)	0.172*** (0.053)	0.206** (0.088)
High sch.	0.420** (0.184)	0.520*** (0.091)	0.571*** (0.111)	0.507*** (0.089)	0.526*** (0.081)	0.548*** (0.158)	0.440*** (0.077)	0.426*** (0.061)	0.424*** (0.063)	0.459*** (0.103)
Vocational sch.	0.399** (0.184)	0.484*** (0.090)	0.607*** (0.113)	0.543*** (0.090)	0.590*** (0.086)	0.616*** (0.164)	0.463*** (0.072)	0.438*** (0.059)	0.338*** (0.059)	0.359*** (0.102)
University	0.939*** (0.185)	1.033*** (0.086)	1.127*** (0.112)	1.123*** (0.092)	1.281*** (0.087)	0.746*** (0.222)	0.733*** (0.096)	0.677*** (0.070)	0.758*** (0.150)	1.074*** (0.169)
Agricul/mining	-0.151 (0.122)	-0.274*** (0.081)	-0.257*** (0.082)	-0.114 (0.108)	-0.084 (0.123)	-0.408*** (0.103)	-0.376*** (0.082)	-0.340*** (0.056)	-0.314*** (0.055)	-0.473*** (0.089)
Manufacture/enr	-0.216*** (0.036)	-0.255*** (0.034)	-0.318*** (0.031)	-0.256*** (0.028)	-0.212*** (0.038)	-0.213*** (0.078)	-0.082 (0.053)	-0.027 (0.035)	-0.181*** (0.042)	-0.328*** (0.079)
Construct	-0.196*** (0.065)	-0.262*** (0.063)	-0.321*** (0.060)	-0.263*** (0.066)	-0.165 (0.116)	-0.024 (0.086)	0.002 (0.047)	0.019 (0.037)	0.003 (0.055)	-0.183** (0.072)

Trade	-0.296*** (0.036)	-0.394*** (0.034)	-0.381*** (0.037)	-0.275*** (0.044)	-0.119** (0.053)	-0.082 (0.065)	-0.064 (0.039)	-0.091*** (0.035)	-0.206*** (0.043)	-0.380*** (0.073)
Transport	-0.391*** (0.080)	-0.395*** (0.064)	-0.466*** (0.042)	-0.421*** (0.057)	-0.346*** (0.103)	-0.196* (0.110)	-0.183*** (0.059)	-0.113** (0.047)	-0.247*** (0.057)	-0.362*** (0.101)
Urban	0.034 (0.038)	0.068** (0.030)	0.043* (0.025)	0.062** (0.027)	0.051 (0.038)	0.108* (0.062)	0.119*** (0.038)	0.148*** (0.027)	0.123*** (0.031)	0.116** (0.057)
Constant	-0.104 (0.194)	0.208** (0.101)	0.503*** (0.121)	0.825*** (0.097)	0.934*** (0.100)	-1.339*** (0.194)	-0.583*** (0.095)	-0.084 (0.067)	0.364*** (0.069)	0.740*** (0.126)
Observations	3,879	3,879	3,879	3,879	3,879	3,448	3,448	3,448	3,448	3,448

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1